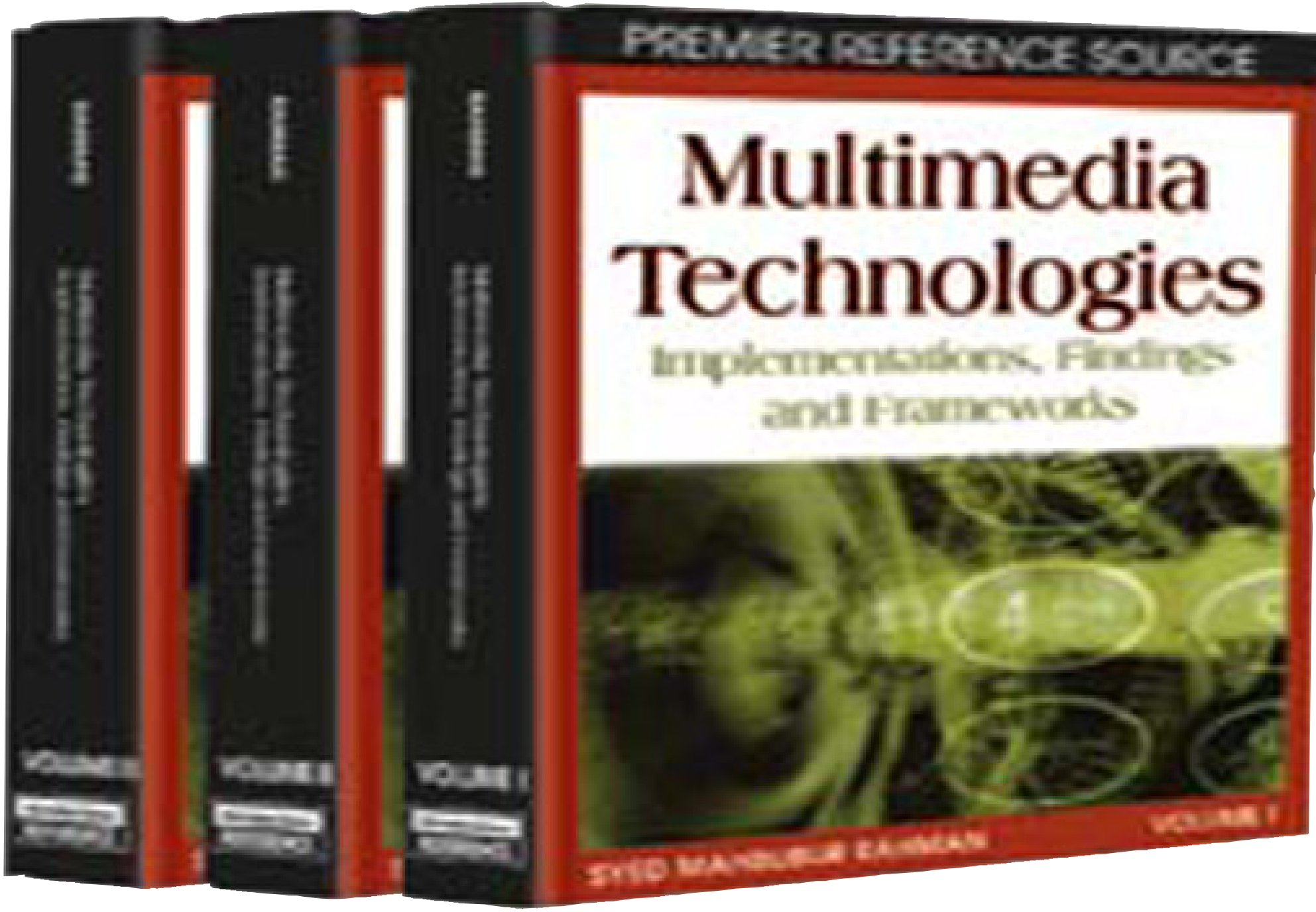

PREMIER REFERENCE SOURCE
Multimedia Technologies
Implementations, Findings and Frameworks
VOLUME I

Multimedia Technologies:
Concepts, Methodologies, Tools, and Applications

Syed Mahbubur Rahman
Minnesota State University, Mankato, USA

Volume I

INFORMATION SCIENCE REFERENCE
Hershey · New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Jamie Snavely
Typesetter: Michael Brehm, Jeff Ash, Carole Coulson, Elizabeth Duke, Chris Hrobak, Sean Woznicki
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Multimedia technologies : concepts, methodologies, tools, and applications / Syed Mahbubur Rahman, editor.
p. cm.
Includes bibliographical references and index.
Summary: "This book offers an in-depth explanation of multimedia technologies within their many specific application areas as well as presenting developing trends for the future"--Provided by publisher.
ISBN 978-1-59904-953-3 (hardcover) -- ISBN 978-1-59904-954-0 (ebook)
1. Multimedia systems. 2. Multimedia communications. I. Syed, Mahbubur Rahman, 1952-
QA76.575.M5218 2008
006.7--dc22
2008021157

Editor-in-Chief

Associate Editors

Editorial Advisory Board

List of Contributors

Contents
by Volume

Volume I

Section 1. Fundamental Concepts and Theories

This section serves as the foundation for this exhaustive reference tool by addressing crucial theories essential to the understanding of multimedia technologies. Chapters found within these pages provide an excellent framework in which to position multimedia technologies within the field of information science and technology. Individual contributions provide overviews of multimedia education, multimedia messaging, and multimedia databases while also exploring critical stumbling blocks of this field. Within this introductory section, the reader can learn and choose from a compendium of expert research on the elemental theories underscoring the research and application of multimedia technologies.

Section 2. Development and Design Methodologies

This section provides in-depth coverage of conceptual architectures, frameworks and methodologies related to the design and implementation of multimedia technologies. Throughout these contributions, research fundamentals in the discipline are presented and discussed. From broad examinations to specific discussions on electronic tools, the research found within this section spans the discipline while also offering detailed, specific discussions. Basic designs, as well as abstract developments, are explained within these chapters, and frameworks for designing successful multimedia interfaces, applications, and even environments are discussed.

Volume II

Section 3. Tools and Technologies

This section presents extensive coverage of the interaction between multimedia and the various tools and technologies that researchers, practitioners, and students alike can implement in their daily lives. These chapters provide an in-depth analysis of mobile multimedia, while also providing insight into new and upcoming technologies, theories, and instruments that will soon be commonplace. Within these rigorously researched chapters, readers are presented with countless examples of the tools that facilitate the transmission of multimedia data. In addition, the successful implementation and resulting impact of these various tools and technologies are discussed within this collection of chapters.

Section 4. Utilization and Application

This section introduces and discusses a variety of the existing applications of multimedia technologies that have influenced education, science, and even music and proposes new ways in which multimedia technologies can be implemented within organizations and in society as a whole. Within these selections, particular multimedia applications, such as face recognition technology and educational, are explored and debated. Contributions included in this section provide excellent coverage of today's multimedia environment and insight into how multimedia technologies impact the fabric of our present-day global village.

Section 5. Organizational and Social Implications

This section includes a wide range of research pertaining to the social and organizational impact of multimedia technologies around the world. Chapters introducing this section analyze multimedia as a vehicle for cultural transmission and language, while later contributions offer an extensive analysis of educational multimedia. The inquiries and methods presented in this section offer insight into the integration of multimedia technologies in social and organizational settings while also emphasizing potential areas of study within the discipline.

Volume III

Section 6. Managerial Impact

This section presents contemporary coverage of the more formal implications of multimedia technologies, more specifically related to the corporate and managerial impact of the core concepts of multimedia, and how these concepts can be applied within organizations. The design and implementation of multimedia advertising is the focus of this section, which provides a how-to guide for multimedia business and commerce. The managerial research provided in this section allows executives and employees alike to understand the role of multimedia technology in business.

Section 7. Critical Issues

This section addresses conceptual and theoretical issues related to the field of multimedia technologies, which include quality of service issues in multimedia transmission and the numerous approaches adopted by researchers that aid in making multimedia technologies more effective. Within these chapters, the reader is presented with analysis of the most current and relevant conceptual inquires within this growing field of study. Particular chapters address methodologies for the organization of multimedia objects and the relationship between cognition and multimedia. Overall, contributions within this section ask unique, often theoretical questions related to the study of multimedia technologies and, more often than not, conclude that solutions are both numerous and contradictory.

Section 8. Emerging Trends

This section highlights research potential within the field of multimedia technologies while exploring uncharted areas of study for the advancement of the discipline. Introducing this section are selections addressing the need for universal access to multimedia. Additional selections discuss the future of multimedia education, advances in multimedia transmission on the Internet, and the possibilities and limitations of multimedia content protection. These contributions, which conclude this exhaustive, multi-volume set, provide emerging trends and suggestions for future research within this rapidly expanding discipline.

Preface

As multimedia technologies and their applications have witnessed explosive growth within the past two decades, information has become increasingly interactive and multidimensional. Traditional text-based data has been augmented and, in some cases, replaced by audiovisual content that is used to transform teaching styles, enhance business transactions, and promote cultural literacy. Researchers, students, and educators have benefited from and been challenged by increased access to multimedia data on the Internet, television, and even on their personal mobile devices. These technologies and their applications will continue to pervade and simplify our daily lives and, as a result, we must continue to understand, develop, and utilize the latest in multimedia research and exploration.

As the study of multimedia technologies and their applications has grown in both number and popularity, researchers and educators have devised a variety of techniques and methodologies to develop, deliver, and, at the same time, evaluate the effectiveness of their use. The explosion of methodologies in the field has created an abundance of new, state-of-the-art literature related to all aspects of this expanding discipline. This body of work allows researchers to learn about the fundamental theories, latest discoveries, and forthcoming trends in the field of multimedia technologies.

Constant technological and theoretical innovation challenges researchers to stay abreast of and continue to develop and deliver methodologies and techniques utilizing the discipline's latest advancements. In order to provide the most comprehensive, in-depth, and current coverage of all related topics and their applications, as well as to offer a single reference source on all conceptual, methodological, technical, and managerial issues in multimedia technology, Information Science Reference is pleased to offer a three-volume reference collection on this rapidly growing discipline. This collection aims to empower researchers, students, and practitioners by facilitating their comprehensive understanding of the most critical areas within this field of study.

This collection, entitled **Multimedia Technologies: Concepts, Methodologies, Tools, and Applications**, is organized into eight distinct sections which are as follows: 1) Fundamental Concepts and Theories, 2) Development and Design Methodologies, 3) Tools and Technologies, 4) Utilization and Application, 5) Organizational and Social Implications, 6) Managerial Impact, 7) Critical Issues, and 8) Emerging Trends. The following paragraphs provide a summary of what is covered in each section of this multi-volume reference collection.

Section One, **Fundamental Concepts and Theories**, serves as a foundation for this exhaustive reference tool by addressing crucial theories essential to understanding multimedia technologies. Opening this elemental section is "Fundamentals of Multimedia" by Palmer W. Agnew and Anne S. Kellerman, which defines the term "multimedia," provides an overview of end-user multimedia devices, and addresses some of the main issues and challenges within the field. Specific applications of multimedia technologies are discussed in selections such as "Core Principles of Educational Multimedia" by Geraldine Torrisi-Steele and "Introduction to Mobile Multimedia Communications" by Gour C. Karmakar, Laurence S.

Dooley, and Michael Mathew. Within the contribution "Multimedia Databases," researcher Mariana Hentea explains the fundamental use and importance of multimedia databases in shaping on-demand television, medical systems, and even fashion design. Similarly, in "Multimedia Information Retrieval at a Crossroad," Qing Li, Jun Yang, and Yueting Zhuang outline the main methods used for retrieving multimedia content and highlight both the current and emerging applications of such technology. The selections within this comprehensive, foundational section enable readers to learn from expert research on the elemental theories underscoring multimedia technologies.

Section Two, **Development and Design Methodologies**, contains in-depth coverage of conceptual architectures and frameworks, providing the reader with a comprehensive understanding of emerging theoretical and conceptual developments within the development and utilization of multimedia technologies. "Content-Based Multimedia Retrieval" by Chia-Hung Wei and Chang-Tsun Li suggests that content-based retrieval of multimedia content, as opposed to traditional text-based retrieval, makes database searching more efficient. Other selections, such as "Cognitively Informed Multimedia Interface Design" by Eshaa M. Alkhalifa, explain how cognitive psychology has helped to change the design of multimedia educational systems. The design of multimedia systems on mobile devices is explored at length in selections such as "Enabling Multimedia Applications in Memory-Limited Mobile Devices" by Raul Fernandes Herbster, Hyggo Almeida, Angelo Perkusich, and Marcos Morais; "Mobile Multimedia Collaborative Services" by Do Van Thanh, Ivar Jørstad, and Schahram Dustdar; and "Design of an Enhanced 3G-Based Mobile Healthcare System" by José Ruiz Mas, Eduardo Antonio Viruete Navarro, Carolina Hernández Ramos, Álvaro Alesanco Iglesias, Julián Fernández Navajas, Antonio Valdovinos Bardají, Robert S. H. Istepanian, and José García Moros. From basic designs to abstract development, chapters such as "Designing for Learning in Narrative Multimedia Environments" by Lisa Gjedde and "On a Design of SCORM-Compliant SMIL-Enabled Multimedia Streaming E-Learning System" by Sheng-Tun Li, Chu-Hung Lin, and Pao-Ta Yu serve to expand the reaches of development and design methodologies within the field of multimedia technologies.

Section Three, **Tools and Technologies**, presents extensive coverage of various tools and technologies and their use in creating and expanding the reaches of multimedia applications. The emergence of mobile devices and the potential for enabling multimedia content on these devices is the subject of articles such as "Multimedia for Mobile Devices" by Kevin Curran; "Multimedia Contents for Mobile Entertainment" by Hong Yan, Lara Wang, and Yang Ye; and "Multimedia Information Design for Mobile Devices" by Mohamed Ally. The new, multimedia-enabled face of distance education is explored throughout Hakikur Rahman's pair of contributions, "Interactive Multimedia Technologies for Distance Education Systems" and "Interactive Multimedia Technologies for Distance Education in Developing Countries." Throughout these selections, Rahman explains how multimedia technologies have transformed the physical classroom into the virtual classroom. Other technologies that are investigated within this section include digital watermarking, in the selections "Digital Watermarking for Multimedia Transaction Tracking" by Dan Yu and Farook Sattar and "Digital Watermarking Schemes for Multimedia Authentication" by Chang-Tsun Li, and communication systems on digital televisions, which are detailed in "Multimedia Communication Services on Digital TV Platforms" by Zbigniew Hulicki. These rigorously researched chapters provide readers with countless examples of the up-and-coming tools and technologies that emerge from or can be applied to the multidimensional field of multimedia technologies.

Section Four, **Utilization and Application**, explores the ways in which multimedia technologies and their applications have been practically employed in a variety of contexts. This collection of innovative research begins with "Integrated-Services Architecture for Internet Multimedia Applications" by Zhonghua Yang, Yanyan Yang, Yaolin Gu, and Robert Gay, which documents the emergence of multimedia applications on the Internet. As music retrieval has been an essential part of the multimedia revolution,

several selections, including "Interactive Multimedia MUSICNETWORK: An Introduction" by Kia Ng and Paolo Nesi and "Content-Based Music Summarization and Classification" by Changsheng Xu, Xi Shao, Namunu C. Maddage, Jesse S. Jin, and Qi Tian, document the classification and study of music-related multimedia technologies. Another application of multimedia technology, face recognition, is studied in "Face Recognition Technology: A Biometric Solution to Security Problems" by Sanjay K. Singh, Mayank Vatsa, Richa Singh, K. K. Shukla, and Lokesh R. Boregowda. As this section concludes, some of the more novel applications of multimedia technologies are surveyed. "Location-Based Multimedia Services for Tourists" by Panagiotis Kalliaras, Athanasios-Dimitrios Sotiriou, P. Papageorgiou, and S. Zoi presents a particular system that aims to provide its users with personalized, tourism-related multimedia information. From established applications to forthcoming innovations, contributions in this section provide excellent coverage of today's global community and demonstrate how multimedia technologies impact the social, economic, and political fabric of our present-day global village.

Section Five, **Organizational and Social Implications**, includes a wide range of research pertaining to the organizational and cultural implications of multimedia technologies. Introducing this section is "Multimedia as a Cross-Channel for Cultures and Languages" by Ramesh C. Sharma and Sanjaya Mishra, a selection that identifies multimedia's role in both preserving and transmitting culture. One of the most widely researched topics in multimedia research—the impact of multimedia technologies upon the modern-day educational system—is explored at length in chapters such as "Teaching, Learning and Multimedia" by Loreen Marie Butcher-Powell, "Planning for Multimedia Learning" by Patrick J. Fahy, and "Web-Based Synchronized Multimedia Lecturing" by Kuo-Yu Liu and Herng-Yow Chen. Other contributions, such as "Web-Based Multimedia Children's Art Cultivation" by Hao-Tung Lin and Herng-Yow Chen and "Student-Generated Multimedia" by Mathew Mitchell, explore the ways in which multimedia authoring among students has been an asset to their overall educational experience. Overall, the discussions presented in this section offer insight into the integration of multimedia technologies into society and the benefit these technologies have provided.

Section Six, **Managerial Impact**, presents contemporary coverage of the applications and implications of multimedia technologies in a business setting. Core concepts such as mobile commerce, Internet security, and the evolution of new advertising techniques are discussed in this collection. "Distanced Leadership and Multimedia" by Stacey L. Connaughton explains how multimedia has allowed virtual teams, who may never see each other face-to-face, communicate more effectively. "Short Message Service (SMS) as an Advertising Medium" by Shintaro Okazaki provides a summary of the use of both SMS and MMS for mobile advertising. Similarly, the selection "Mobile Multimedia for Commerce" by P.M. Melliar-Smith and L.E. Moser describes the rise of mobile commerce and explains that, since mobile devices are portable, they are the perfect tools for convenient selling and purchasing. Tziporah Stern's "Internet Privacy from the Individual and Business Perspectives" explains how online privacy has become a major issue for businesses and individuals alike. Within this selection, Stern provides both an overview of the problem and a list of potential solutions that individuals and businesses can adopt to make Web browsing and shopping more secure. The comprehensive research in this section offers an overview of the major issues that businesses must address in order to remain successful and current in an environment rich with multimedia content and interaction.

Section Seven, **Critical Issues**, presents readers with an in-depth analysis of the more theoretical and conceptual issues within this growing field of study by addressing topics such as quality of service (QoS) and multimedia security. "Distributed Approach for QoS Guarantee to Wireless Multimedia" by Kumar S. Chetan, P. Venkataram, and Ranapratap Sircar and "Quality of Service Issues in Mobile Multimedia Transmission" by Nalin Sharda discuss current and potential challenges in successfully transmitting multimedia content over wireless networks. Daniel J. Buehrer's "Organizing Multimedia

Objects by Using Class Algebra" highlights a particular method for retrieving Web-based multimedia content. Similarly, specific methods for image retrieval and indexing are presented and defined in selections such as "A Stochastic and Content-Based Image Retrieval Mechanism" by Mei-Ling Shyu, Shu-Ching Chen, and Chengcui Zhang and "A Spatial Relationship Method Supports Image Indexing and Similarity Retrieval" by Ying-Hong Wang. Concluding this collection of the more conceptual issues within the field is "Multimedia Security and Digital Rights Management Technology" by Eduardo Fernandez-Medina, Sabrina De Capitani di Vimercati, Ernesto Damiani, Mario Piattini, and Pierangela Samarati, in which the importance of defining who has the rights to particular multimedia content is addressed. In all, the theoretical and abstract issues presented and analyzed within this collection form the backbone of revolutionary research in multimedia technologies and their applications.

The concluding section of this authoritative reference tool, **Emerging Trends**, highlights research potential within the field of multimedia technologies while exploring uncharted areas of study for the advancement of the discipline. New IETF transport layer protocols in support of multimedia data transmission are presented and explored within Michael Welzl's "New Internet Protocols for Multimedia Transmission," while the potential for a multimedia world is debated within "Universal Multimedia Access" by Andrea Cavallaro. Later selections, such as "Toward Effective Use of Multimedia Technologies in Education" by Geraldine Torrisi-Steele, maintain that the current use of multimedia in education is problematic and present guidelines for the future construction of multimedia-informed educational systems. Similarly, in their contribution "Future Directions of Multimedia Technologies in E-Learning," researchers Timothy K. Shih, Qing Li, and Jason C. Hung argue for the incorporation of pedagogic theory into the design of distance learning systems which will, in their opinion, make learning more efficient. This final section demonstrates that multimedia technologies, with their infinite potential for application, will continue to both shape and define the way we access and absorb information.

Although the contents of this multi-volume book are organized within the preceding eight sections which offer a progression of coverage of the important concepts, methodologies, technologies, applications, social issues, and emerging trends, the reader can also identify specific contents by utilizing the extensive indexing system listed at the end of each volume. Furthermore, to ensure that the scholar, researcher, and educator have access to the entire contents of this multi-volume set, as well as additional coverage that could not be included in the print version of this publication, the publisher will provide unlimited, multi-user electronic access to the online aggregated database of this collection for the life of the edition, free of charge when a library purchases a print copy. In addition to providing content not included within the print version, this aggregated database is also continually updated to ensure that the most current research is available to those interested in multimedia technologies.

Within the past two decades, multimedia content began to occupy our computers, television screens, and mobile devices and has made the information age a reality. The applications of multimedia technology are both diverse and innumerable—education, business, and even art have been revolutionized by an easily accessible flood of interactive data. Research into how to properly access, distribute, and guarantee the security of multimedia content has implications in constructing multimedia interfaces, ensuring that mobile devices effectively transmit multimedia data, and shaping the future of commerce. With continued innovation in multimedia technologies and their applications and ongoing research into the best ways to distribute and utilize multimedia content, the discipline will continue to grow and transform as our world becomes more interactive.

The diverse and comprehensive coverage of multimedia technologies in this three-volume, authoritative publication will contribute to a better understanding of all topics, research, and discoveries in this developing, significant field of study. Furthermore, the contributions included in this multi-volume collection series will be instrumental in the expansion of the body of knowledge in this enormous field,

resulting in a greater understanding of the fundamentals while also fueling the research initiatives in emerging fields. We at Information Science Reference, along with the editor of this collection, hope that this multi-volume collection will become instrumental in the expansion of the discipline and will promote the continued growth of multimedia technologies.

Introductory Chapter
A Brief Introduction to the Field of Multimedia Technologies

Syed Mahbubur Rahman
Minnesota State University, Mankato, USA

INTRODUCTION

In an age where information rules, computer-based multimedia technology is a tool for communicators of all trades and is also an effective catalyst for change. Multimedia is a technology that allows us to present text, audio, images, animations, and video in an interactive way that has created a tremendous impact on all aspects of our day-to-day life. It also has the potential to continue to create ever more fascinating applications, some of which are described in articles listed in the references.

Technology is changing the world. However, the technology by itself can not change the world. It is the people who adopt and use the technology that make the changes. Accordingly, the inherent property of multimedia to support human-centered computing (HCC) may be credited for its explosive growth in all areas of application as is evident from several research works included in the references. Becvar has investigated how video blogging ("vlogging") systems affect the learning practices involved in training novice practitioners, and how integrating new technology alters the complex social dynamics of professional training (Becvar, 2007).

Present-day information access, which is a part of our everyday life, invariably involves multimedia data in some form or another. The use of multimedia enhances a user's ability to communicate and collaborate. A major component required in multimedia applications is a computer with high processing speed and large storage capacity. The hardware cost is decreasing at a rate never seen before, along with a rapid increase in the storage capacity, computing power, and network bandwidth. These developments have made use of multimedia more affordable and contributed to the tremendous growth in production and use of multimedia contents that we are experiencing recently. Multimedia technology has demonstrated the potential to evolve the paradigm of end-user computing, from interactive text and graphics model, into one more compatible with the digital electronic world of the 21st century. It is almost impossible to track the magnitude and breadth of the changes that multimedia and communication technology is undergoing.

This is an introductory chapter which aims to present basic terminology, tools, formats, and content protection used in multimedia applications and their development. It discusses basics of multimedia networking, security, issues, and trends. It also includes an extensive list of references on latest research issues and future trends on multimedia development and application areas.

WHAT IS MULTIMEDIA?

The word **multimedia**, originating from the Latin words "multum" and "medium", means combination of multiple media contents. It is a technology that, even today after two decades of explosive growth, means different things to different people. It might be an artistic medium or a communication tool or a teaching and learning tool for some, while it might be a way to complete a business transaction for others. In general, multimedia includes a combination of text, audio, still images, animation, video, and interactive content. The integration of multimedia technology into the communication environment has the potential to transform an audience from passive recipients of information to active participants in a media-rich learning process. The term "rich media" is synonymous for interactive multimedia. Currently multimedia is widely used as a computer-based interactive communications process that includes any combination of static (text, graphics, and still images, etc.) and active (sound, animation, and video, etc.) media. Inclusion of more than one media with at least one of them as active media is required to preserve the definition of multimedia. Most current day Web pages are examples of use of multimedia. Multimedia provides a real world feeling by incorporating a multi-sensory experience.

EVOLUTION OF MULTIMEDIA

Throughout the history of digital technologies, multimedia has always existed in one form or another. Not too long ago multimedia was defined as the combination of images with sound. A common implementation of early "multimedia" was refrigerator-size kiosks containing both a monitor and laserdisc player hooked up to a push-to-start button. The multimedia presentations were linear in early days. The users were presented with information to listen or read or do both as passive witness without any form of interaction, except to push the start button. In other early implementations of multimedia, for example, a narrator would tell a story with a series of still pictures or even videos filling the screen. There were no ways a viewer could randomly access the specific sections of the presentation which were more important or useful and skip the boring parts. Users were not very impressed with this form of technology and were looking for innovative approaches. With technological development, interactivity has been introduced with videos being played from a laserdisc controlled by some computer that make use of the random-access capability of the laserdisc player and allow a viewer to select random modules of content. This capability empowered the users to control the path and flow of content, while bypassing superfluous topics. With hyperlink and hypermedia interface a user can navigate through into the subject areas of his/her interest.

As new developments are taking place, the multimedia technology is supported with increased hard disk capacity, higher data bandwidth, new video compression technologies, more sophisticated multimedia software, and CD/DVD included as a common feature of computers. Within a short span of time, personal computers proved to be an appropriate medium to deliver full screen, full motion video with their innovative technologies and video compression technologies. Apple computer was one of the early computers coupled with these multimedia features.

The definition of multimedia has now been restructured—a mix of elements of hyperlinked text, animation, graphics, video, and audio, in an interactive environment. This technology is constantly changing and evolving as new product versions arrive each day replacing the older ones that has resulted a competition for high-tech corporations in which only the fittest and the best can survive.

A good multimedia application is one that keeps the technology invisible from the user. The purpose of a good multimedia presentation is to envelope the viewer with rich text, clear sound, sharp image, and smooth motion that can be stopped, started, and cross-referenced with ease.

CATEGORIES OF MULTIMEDIA

Multimedia may be divided into following three categories based on their functions and how they are organized.

- Linear and non-linear;
- Interactive and non-interactive; and
- Real-time and recorded.

Linear active content progresses without any navigation control for the viewer such as a cinema presentation. *Non-linear* content offers user interactivity to control progress as used with a computer game or used in self-paced computer-based training.

Interactive multimedia is the means to interface with different media through input (e.g., a computer keyboard, mouse, touch screen, on screen buttons, and text entry, etc.) and output devices allowing a user to make decisions as to what takes place next with multimedia.

Multimedia presentations can be *live* (real-time) or *recorded*. A recorded presentation may allow interactivity via a navigation system. A live multimedia presentation may allow interactivity via interaction with the presenter or performer. More details are available in the reference section.

MULTIMEDIA FILE FORMATS

Multimedia formats represent the various ways each media type is stored and used to transport multimedia data. This section covers a range of multimedia file formats—especially audio, video, and image file formats—used in the delivery of multimedia and includes both formal standards and de facto standards. Media formats may differ for the purpose it is used, for example, for streaming or for downloading.

Formats suitable for *streaming* should be able to transfer data in a continuous stream (usually for audio or video), for example, over the Internet, so that the player can play it as the data arrives without waiting for the entire file to be downloaded. At start the streaming media players store several seconds worth of data, known as a buffer, in its memory. The player then begins playback of the file from the buffer, while the data continues to fill in. The buffer, as the name suggests, allows for continuous playback of the audio or video by compensating for any delays in the transmission of the rest of the file. The buffer absorbs the bursts of data as they are received and releases it at a constant bit rate for smooth playback. For *downloading* purpose, the file format does not need to satisfy this criterion of transferring data continuously.

There are several formats for each media with different features requiring the players to be capable of decoding these features. Fortunately many hardware and software players are now available, which are able to support common multiple formats. There are also tools available for converting a media in one format to different formats (Pereira, 2007; Viljoen, Calitz, & Cowley, 2003).

Audio Formats: There are a number of different *types* of audio file formats. The most common are wave files (wav), MPEG Layer-3 files (mp3) and AIFF (Audio Interchange File Format) file types. Common streaming audio formats include Windows Media Audio/Active Streaming Format (ASF), QuickTime, RealAudio, and MP3. There are, however, some other audio file types included in the Table 1. The format types are also represented by the file extension. Codec determines the way the audio is compressed and stored. Some file types always use a particular codec, while others may support multiple codecs.

Initial internet browsers, such as Netscape Navigator and Mosaic, included AU (Audio, or m-law), WAV (Waveform Audio), and AIFF (Audio Interchange File Format) formats and got widespread Internet acceptance. These formats are considered as better choices for inclusion on web pages because of their cross-platform support, and their ability to run natively within a browser. Both WAV and AIFF are commonly used on the Web at lower-quality sampling rates to keep file size reasonable and reduce upload and download time.

Most of the streaming multimedia formats are proprietary and require special servers for encoding and transmission. Some commonly used streaming audio formats are RealAudio, XingMPEG, and streaming audio for the popular Shockwave plug-in. XingMPEG and Shockwave each have compression ratio of up to 26:1 and 176:1, respectively. For both these formats the average file size varies depending on the strength of compression. The two streaming formats, MPEG (Moving Pictures Expert Group) and RealAudio have changed audio on the Internet. MPEG is best known as a standard for the compression of video files claiming very high compression ratios as high as 26:1. However, ratios greater than 12:1 are seldom used due to the sacrifices in quality. Even at this compression ratio MPEG is capable of delivering CD-quality audio files that in any other format is extremely difficult to achieve. RealAudio has also demonstrated itself to be one of the fastest growing proprietary streaming formats. RealAudio Player allows users to download both on-demand and live-broadcast streaming audio files. The RealAudio file

Table 1. Common digital audio formats

Audio Formats	Extensions	Codec	Characteristics
AU (Sun/Next)	.au	*u-law	- Relatively good compression and small file size (2:1 and 8 Kb/sec). - 8 bit encoding only; sound acceptable, but not premium quality. - introduced by Sun Microsystems and NeXT Compute.
WAV	.wav	*PCM	- 16 bit, better sound quality than AU. - 1 minute audio consume more than 10MB. - Native support in windows. - No compression and big file size (10 Mb/s). - Higher quality than AU.
AIFF (Mac)	.aif, .aiff	*PCM	- WAV comparable sound quality. - 8 or 16 bit sampling. - Native support on Mac. - Small file size with 8 bit sampling.
MPEG audio	.mp2 .aac	MPEG Audio	- Good sound quality. - Sometimes proprietary nature makes them incompatible.
MP3	.mp3	MPEG Audio Layer-III	- compressed to 10:1 of an equivalent PCM. - most recommended for music.
Windows Media Audio	.wma	Proprietary (Microsoft)	- Designed with Digital Rights Management (DRM) abilities for copy protection.
QuickTime	.qt	Proprietary (Apple Computer)	- supports both streaming audio and streaming video. - widely used for streaming video on the Web.
RealAudio	.ra, ram	Proprietary (Real Networks)	- supports both streaming audio and streaming video.

** Can be used with other codecs*

format supports two levels of compression: an "AM sound" quality for 14.4-Kbps modems and an "FM sound" quality for 28.8-Kbps and faster connections. However, audio often "gaps out" on 14.4-28.8 modem connections. Depending on the algorithm used, RealAudio files typically require only 1.1 to 2.4 kilobytes per second or between 3.6 MB and 8 MB per hour of audio. MetaVoice (VOX) format is a good choice for delivering speech.

Video Formats: Some common video formats are AVI (Audio Video Interleave), MPEG (Moving Picture Expert Group), and QuickTime (MOV). A breakdown of these formats is shown in Table 2.

Because movies involve both audio and video, the file sizes as a rule are very large. QuickTime supports compression ratios as high as 50:1, yet a typical MOV file consumes almost 70 KB per second of video, or about 4 MB per minute. MPEG compression ratios can be as big as 200:1. Accordingly, MPEG video files are usually smaller than QuickTime movies. Because of these advantages, MPEG format is a viable alternative for inclusion of video in internet based applications. AVI's widespread acceptance is due to native support by the Windows Media Player application.

Some common streaming video formats include XingMPEG video and VDOnet (VDO) file types. There are codecs available that enable WAV, MPEG, QT, and standard AVI streaming. Due to the large amount of data transfer that video demands, only those users connecting at T1 speeds may be able to properly view a full-motion video. VDO uses a "wavelet" method to compress AVI files into its proprietary format. Although compression ratios vary, VDO files average about 1 MB per minute of video.

H.264: H.264/Advanced Video Coding (AVC) is a new video compression standard developed jointly by the Joint Video Team of ITU-T Q.6/SG16 VCEG and ISO/IEC MPEG. The partnership effort is also known as Joint Video Team (JVT). The final drafting work on the first version of the standard was completed in May 2003. This relatively new standard has already gained wide industry acceptance and is being adopted in many regional and industry-specific standards because of its high compression efficiency when compared to similar schemes such as H.262/MPEG-2. So it deserves more detailed discussion compared to the other standards.

This standard is known by different names, such as, H.264/AVC or AVC/H.264 or H.264/MPEG-4 AVC or MPEG-4/H.264 AVC. H.264 is a name related to the ITU-T line of H.26x video standards. AVC relates to the ISO/IEC MPEG side of the partnership project. H.264 is also referred to as "MPEG-4 Part 10" (part of the MPEG-4 specification, formally, ISO/IEC 14496-10). The ITU-T H.264 standard and the ISO/IEC MPEG-4 Part 10 standard are jointly maintained so that they have identical technical content.

The H.264/AVC standard provides much better video quality (reduction in artifacts such as blockiness, color bands, etc.) at substantially lower bit rates, higher resolution and with lower storage requirements compared to the previous compression schemes. It also allows flexibility to be applied to a very wide variety of applications (e.g., for both low and high bit rates, and low and high resolution video) and work well on wide variety of networks and systems (e.g., for broadcast, DVD storage, RTP/IP packet networks, and ITU-T multimedia telephony systems).

H.264 is becoming the worldwide digital video standard for consumer electronics and personal computers. H.264 has been adopted by the Motion Picture Experts Group (MPEG) to be a key video compression scheme in the MPEG-4 format for digital media exchange. Following are some important applications where the standard has been adopted.

- Broadcast over cable, satellite, cable modem, DSL, terrestrial, and so forth. Examples include:
  - Broadcast television in Europe (approved by the Digital Video Broadcast (DVB) standards body in Europe in late 2004).
  - Receivers of HDTV and pay TV channels for digital terrestrial broadcast television services (referred to as "TNT") in France in late 2004.

Table 2. Common video formats

Video Formats	Characteristics
AVI (Audio Video Interleave)	• Uncompressed results high quality video, but large files. • Often there is problem synchronizing audio with video. • The entire file must be downloaded before being played. • Native support on Windows. • Average AVI file size is between a MOV file size and MPEG file size.
MPEG (Motion Pictures Experts Group)	• Can provide VHS or better quality movies. • Up to 200:1 compression ratio and storage rate is 2.8MB/minute. • Can produce full-motion video with relatively small file size. • Typically MPEG1 is used to make a one hour VCD movie. • MPEG1 quality is same as VHS. • MPEG2 quality is better than VHS and used to make DVD video, can be used to make about a 30 minute high quality VCD.
MPEG - 4	• ISO/IEC standard developed in 1998 by MPEG. • Includes many features of MPEG-1 and MPEG-2 and other related standards. • New features include extended VRML support for 3D rendering, object-oriented composite files (including audio, video and VRML objects). • Multiplexes and synchronizes data, associated with media objects, in such a way that they could be transported further via network channels. • Developed primarily for low bit-rate video communications and later its scope was expanded to be much more of a multimedia coding standard. • Enables developers to control their content better and to fight more effectively against copyright violations. • MPEG-4 part 10 (MPEG-4 AVC/H.264) is becoming a more accepted standard. (H.264 is discussed below as a separate section).
MOV (Apple QuickTime Movie)	• Requires Apples QuickTime Movie Player. • Runs natively on the Mac platform. • Up to 50:1 compression ratio and storage rate is 4 MB/minute. • Relatively large file size. • Depending on Compression chosen it can provide a very high quality video clip. However, better quality video requires more storage space. • Can be streamed across the Internet and viewed before entire file has been downloaded using a QuickTime streaming server.
ASF (Advanced Systems Format) WMV (Windows Media Video)	• Can be streamed across the Internet and viewed before entire file has been downloaded when using a Windows Media server. • Audio and/or Video content can be compressed with a wide variety of codecs. • An extensible file format designed to store synchronized multimedia data. • Requires Windows Media Player be installed on client.
RM (Real Media)	• Can be streamed across the Internet and viewed before entire file has been downloaded when using a Real Networks Streaming server. • Has very high compression, but at a cost to quality. • Requires Real Networks RealPlayer to view content.

- The Digital Multimedia Broadcast (DMB) service in the Republic of Korea.
- Mobile-segment terrestrial broadcast services of ISDB-T in Japan.
- Major broadcasters in Japan including NHK, Tokyo Broadcasting System (TBS), Nippon Television (NTV), TV Asahi, Fuji Television, TV Tokyo.
- Direct broadcast satellite TV services such as DirecTV, Dish Network in U.S.; Euro1080 in Europe; Premiere, ProSieben HD & Sat1 HD in Germany; BSkyB in the United Kingdom and Ireland, etc.

- Interactive or serial storage on optical and magnetic devices, DVD, etc. For example, it has been selected as a key compression scheme (codec) for the next generation of optical disc formats, HD-DVD format of DVD forum and Blu-ray disc (sometimes referred to as BD or BD-ROM) format of the Blu-ray Disc Association.
- Conversational services over ISDN, Ethernet, LAN, DSL, wireless and mobile networks, modems, etc., or mixtures of these.
- Video-on-demand or multimedia streaming services over ISDN, cable modem, DSL, LAN, wireless networks, etc.
- Multimedia messaging services (MMS) over ISDN, DSL, Ethernet, LAN, wireless and mobile networks, etc.
- Quick Time, Flash Player, YouTube
- In most application areas of the Motion Imagery Standards Board (MISB) of the United States Department of Defense (DoD).
- For international military use by the North Atlantic Treaty Organization (NATO).

Several companies are producing custom chips capable of decoding H.264/AVC video, will allow widespread deployment of low-cost devices capable of playing H.264/AVC video at standard-definition and high-definition television resolutions. More details can be found at http://en.wikipedia.org/wiki/H.264.

Image File Format: Most common image file formats include Joint Photographic Experts Group (JPEG - .jpg), PDF (use Type I postscript fonts - .pdf), GIF (.gif), TIFF, and so forth. Table 3 illustrates some commonly used file formats and their major features.

MULTIMEDIA TOOLS

Multimedia tools include hardware and software used in the process of developing multimedia applications, for delivering multimedia products and for later maintenance and modifications/update of the products. Most commonly used multimedia hardware includes video, sound cards, player, and recording devices. The hardware devices are going through dramatic improvements almost on a daily basis in their features and ease of use. Multimedia software development tools may be divided according to nature of their use as described below.

1. *Classification of multimedia tools based on their use in different phases of applications development*.
   - *Analysis Tools*: These tools help the designer to study existing system (if any) to identify strength, weaknesses, and opportunities for improvement and analyze the needs and goals for the project scope.

Table 3. Common image file format

Image File formats	Major Features
BMP	- 24 bit or 16.7 million colors (RGB). - 8-bit palette, or 256 colors (RLE), reduces the file size to about 10:1 ratio compared to BMP-RGB.
GIF	- 8-bit palette or 256 colors. - Supports animation and is still widely used to provide image animation effects.
JPEG	- Developed in 1991. In most cases lossy format. - 24-bit total for red, green, and blue. Produces relatively small file sizes. - Min. compression about 5:1 ratio may be achieved compared to BMP-RGB. - Min. progressive compression about 7:1 ratio may be achieved. - Max. compression about 50:1 ratio may be achieved compared to BMP-RGB. - Max. progressive compression about 70:1 ratio may be achieved. - JPG quality is not preferred for archiving master copies.
JPEG-2000	- Developed to achieve a better image quality in a smaller file. - The standardized filename extension is .jp2 for ISO/IEC 15444-1 conforming files and .jpx for ISO/IEC 15444-2 files. The MIME type is image/jp2. - JPEG 2000 uses wavelet based compression, whereas JPEG used the DCT compression. - JPEG 2000 is not widely supported in web browsers.
PCX	- 24 bit or 16.7 million colors.
TIFF Tagged Image File Format	- image format that normally saves 8, 16, 24 or 48 bits per color. - Gray scale – 8 or 16 bit. - Line Art (bi-level) - 1 bit. - Can be lossy or lossless; provide relatively lossless compression. - Widely used as printing industry photograph standard. - Simple and widely used for good quality archived master.
TIFF-LZW compression	- Internally used in Windows. - Not compressed.
RAW	- Not standardized. - Lossless or nearly lossless compression. - Smaller file size than TIFF.
PNG (Portable Network Graphics)	- RGB - 24 or 48 bits, Grayscale - 8 or 16 bits, Indexed color - 1 to 8 bits, Line Art (bi-level) - 1 bit. - Compress files at a similar ratio to jpeg. Unlike jpeg, resaving an image will not degrade its quality. - Used for good quality archive or master.

- *Design Tools*: The design tools are useful for planning the project development, including user characteristics and specific objectives.
- *Management Tools*: These tools are used for the management of the multimedia development process as a whole.
- *Production Tools*: These tools are helpful in the actual production of the multimedia product.
- *Evaluation Tools*: These tools support the task of generating evaluations by various different means.

2. *Classification of multimedia tools based on the media type for which a software tool is used*.
   - Audio software
   - Graphic software
   - Video software
3. *Classification of multimedia tools based on the function of tool*. Multimedia software tools encompass a wide variety of software that a developer needs to use for different functionalities, for example, capturing, playing, combining images, text, music and sound, animation, video, and other special effects. Following are some software tools based on their functions belonging to each of these media categories (i.e., audio, video, text and graphics)
   - Media editors
   - Media viewers
   - Media recorders
   - Media format converters (e.g. bmp to jpeg)
   - Media converters (e.g. text to speech or speech to text converter etc.)
   - Media capture
   - Animators
   - Movie joiner and splitter
   - Watermarking tool
   - Multimedia for the Web
   - Business presentation tool
   - Screen saver creation
   - Slide show software
   - Multimedia photo albums
   - Multimedia authoring

There are several software tools available that are capable of providing the above functionalities separately and for each different media. However, the most commonly used tools have one or several built in functionalities to support several formats and media. For example, Windows media player plays both audio and video in several formats, such as ASF, Real Video/Real Audio 4.0, MPEG 1, MPEG 2, WAV, AVI, MIDI, MOV, VOD, AU, MP3, and QuickTime files etc. CoolEdit can edit, mix and add sound affect in a variety of audio file formats. Avid SoftImage, Animated Gif building packages such as *GifBuilder* can create and support animation; *animated greetings software* allows the user to create personalized

Table 4. Image/graphics/video editing tools

Tool Name	Major Features
Adobe Photoshop	• Allows layers of images, graphics and text • Includes many graphics drawing and painting tools • Sophisticate lighting effects filter • A good graphics, image processing and manipulation tool
Adobe Premiere	• Provides large number (up to 99) of video and audio tracks, superimpositions and virtual clips • Supports various transitions, filters and motions for clips • A reasonable desktop video editing tool
Macromedia Freehand	• Graphics drawing editing package

Table 5. Examples of additional multimedia tools

Name	Author	Type	First public release date	Operating System	Some Features	Protocols supported	Important formats
iTunes	Apple Inc.	Audio, video	2001	Win, Mac	Vp, os, md, vz	HTTP RTSP Podcasting	Mpeg-1, 2, 4, Flash, wmv, Asf, quicktime
QuickTime	Apple Computer	Audio, video	1991	Win, Mac	Vp, os, md, vz	HTTP RTSP Podcasting	Mpeg-1, 2, 4, flash, Mp3, wma, avi, quicktime, mp4
RealPlayer	RealNetworks	Audio, video	1995	Win, Mac, Linux, other unix, Solaris	Vp, md, vz	HTTP RTSP MMS Podcasting	Mpeg-1, 2, mpeg-4, flash, MP3, WMA, RealAudio Vorbis, avi, asf, quicktime, mp4
Windows Media Player	Microsoft	Audio, video	1992	Win, mac, win mobile	Vp, os, md, vz	HTTP RTSP MMS Podcasting	Most formats
Media Player Classic	Gabest	Audio, video	2003	win	vp	HTTP MMS	Mp3, wma, avi, asf, quicktime, mp4

electronic greetings with text, images, and audio files of different formats according to user's preferences. Table 4 and 5 contains some major features of few well known image/graphics/video tools.

Many other software with similar functionalities is also available commercially and also in the public domain.

Multimedia Authoring: Multimedia authoring tools are by far the most versatile and have lots of interactive controls for the user to develop complete multimedia applications from simple (e.g., slide show presentation) to most complex ones (e.g., computer games or interactive computer aided learning applications). Most authoring tools support WYSIWYG (*what you see is what you get*) environment or in a timeline-based environment. Programming and scripting languages are also supported for designing customized and advanced scenes. Other important feature includes exporting the developed projects to self-executing and self-installing files to a CD or DVD recording media. Some well known authoring tools available are Macromedia Director, Authorware, Flash, Hypercard, Hyper studio and IconAuthor etc. Important features of some authoring tools are documented in Table 6.

There are literally thousands of uses for this type of software, for example, computer-based training, surveys, quizzes and tests, encyclopedias, games, interactive kiosks, interactive presentations, screen savers, CD-ROM/DVD content creation, and advertisements, and so forth (Miguel, Barracel, & Panyella, 2004; Petridis, Saathoff, & Dasiopoulou, 2006).

Delivery of Multimedia Products: To make the multimedia products ready for distribution, most of the software tools have some capability of exporting the developed projects to self-executing, self-installing files. These categories of products have the capability to export contents as screen savers. An important feature to note about export specification is its ability to create cross-platform compatibility, create stand-alone files for a specific platform and/or for Web applications. Most animated greetings software creates self-executable files or Web publishable files.

Table 6. Authoring tools

Authoring Tool	Major Features
Macromedia Director	• A multimedia developing tool for any platform: CD's, DVD's, kiosks and of course, the Web. • Proven authoring tool for creating powerful games, simulations and multimedia applications. • Intuitive design and object-oriented development environment. • Authoring tool for creating powerful games, simulations and multimedia applications. • Support JavaScript, Flash(TM) MX 2004 content, DVD-Video, and the ability to create projector files for both Mac and Windows platforms in one simple step. • Full support for ECMAScript-compliant JavaScript syntax, which supplements traditional Lingo support. • Director's powerful features include two scripting languages, cross-platform publishing, and Flash integration. • combines broad support for media types, ease of use, bit-mapped graphics, true 3D rendering, high-performance, and an infinitely extendible development environment to deliver games, rich content and applications for the Internet, desktop, CDs and DVDs. • Lingo script language with own debugger allows more control including external devices, e.g., VCRs and video disk players. • Movie metaphor (the cast includes bitmapped sprites, scripts, music, sounds, and palettes, etc.). Ready for building more interactivities (buttons, etc.). • Inclusion of Lingo, its object-based scripting language, has made it the animation-capable program. The AfterBurner compression Xtra creates Shockwave files, allowing Web playback.
Authorware	• Professional multimedia authoring tool. • Authorware uses the same JavaScript engine found in Macromedia Dreamweaver MX. • All properties within Authorware can now be scripted making it easier for advanced developers to create commands, Knowledge Objects, and extensible content. • Create dynamic, data-driven applications by importing or exporting web-standard XML files into other applications. • Generate tab navigation and captions, and turn text into speech to comply with accessible software legislation. • Supports interactive applications with hyperlinks, drag-and-drop controls, and integrated animation. • Compatibility between files produced from PC version and MAC version.
Flash	• Developed by: Macromedia. • Platforms: Mac, Windows95, NT, WWW (via Flash Player). • A cast/score/scripting tool. • Primarily uses vector graphics. • Can create vector graphics from imported bitmaps. • Create engaging rich-media advertisements: including webmercials, interstitials, and banners. • XML transfer and HTML text support. • ActionScript development tools. • Produce standalone run-time animation applications.

DEVELOPMENT OF MULTIMEDIA APPLICATIONS

In the process of developing multimedia applications, there are some issues that need to be considered, some of which are listed in the following:

- The target audience
- The objective(s) of the application
- The structure of the application

- Multimedia building blocks used to present the proposed content
- The desired degree of interactivity between the user and the computer
- The expected level of user response

Development Steps: The development process for a multimedia title may be divided into several steps as below.

- *Project Planning*: In this step the development team presents a series of main ideas, facts, or short descriptions. Team decides on the project duration, milestones, and finance.
- *Project Design*: This is a step where the team discusses exactly what the application is going to do and how the interface has to be. Decides on the type of project and degree of non-linearity, interactivity. The presentation may be a series of charts and interfaces as detailed below and are very important in the design process of developing a multimedia project. These ideas and design concepts are used in the proposal or for presentation to a client and also in the production process.

    - Flow Chart shows the information flow of the project in the form of a diagram that should include all menus and submenus, introductory and information screens with text, image, audio and video contents. Charts contain boxes that indicate the various sections, which are connected by lines and arrows to indicate links, hierarchy and flow. Flow indicated by arrows can go from top down, or from left to right. The designed charts are expected go through various stages of modifications. Graphics program such as Photoshop or Illustrator may be used to create the chart.
    - Interface Sketches contain the detailed of each menu showing design concept, placement of titles, typographical styles, order/hierarchy of buttons, return arrows, interactivity, borders, image, graphics, video, and so forth.

- *Story Boarding*: For video material, including interviews, documentary, narrative pieces, and so forth, create a simple storyboard(s) that outlines the sequence of events, dialogue, texts, locations, camera shots, sound, music, and so forth. This provides a rough guide for shooting and editing, and need to be updated during the production process. A blueprint of the application is decided upon and created, including the navigation diagram to move from a story to a different story. Choice of media types plays an important role in deciding the acceptability of the final project.
- *Script Development*: This is where the content of the application is developed that would include narrations and images.
- *Production*: The backgrounds, scenes, transitions, animations and other presentation details are done.
- *Authoring*: In this step the developer puts together everything and gives it the form outlined during the project design stage.

A well-designed multimedia project must have a flow with a feeling of continuity that must draw interest and attention of the user. The following are some other considerations that need to be addressed during development of a successful multimedia project.

- The product should have a title or description.
- Graphics, images, and video are placed in the content ergonomically appropriate to the theme of the project and used only to enhance the intended information.

- Careful considerations must be given in the use of correct colors, video, animation, music, and sound effects to add powerful effects in the project design and make it successful.
- Visual message (image and video) should be used when it is clearer than text alone and when it demonstrates a procedure, process, or role-playing technique that is not well presented with text.
- Videos included should be short. Instead of showing a long video all at once, split it up with text, audio, or questions in between. Make sure that the video adds content to the theme.
- The text content need to be accurate, well written, complete with proper grammar and punctuation.
- The success of a large-scale multimedia project depends on a collective effort made by a team of talented communication and production experts—writers, designers, producers, audio and video technicians, systems engineers, and programmers (Kaskalis, Tzidamis, Margaritis, & Evangelidis, 2006; Miguel, Barracel, & Panyella, 2004; Petridis, Simou, & Dasiopoulou, 2006).

MULTIMEDIA COPYRIGHT PROTECTION

Piracy and other copyright violations regarding digital multimedia content represent a significant problem for legal content owners and content distributors. Preventing illegal content copying and distribution is a very difficult problem to solve. So, the digital rights management (DRM) for multimedia, that is, the protection of intellectual property rights for multimedia content, is an important area of interest.

Two methods, *steganography and cryptography*, are well known for copyright protection.

Steganography is the science of hiding information in such a way that only the owner and the intended recipient knows about the message. Others cannot see any change in the information content and can not even realize that there is a hidden message/owner's identity.

Due to the nature of human vision and hearing on one hand and typical highly redundant properties of multimedia documents on the other hand, a small noise can be inserted into an image, audio and video documents without any visible deterioration of the quality of audio, image, or video. A simple example of hiding data within an image file may be to include data by a method known as least significant bit (LSB) insertion. In this method, the binary representation of the hidden data is written in the image by modifying the LSB of each byte representing the color of each pixel of the image. Due to this modification of a bit in a pixel, the amount of visible change in the image is minimal and indiscernible to the human eye. A digital watermark aims to accomplish a similar function. For example, an image may have an embedded hidden signature so that the owner can later prove his/her ownership in case others attempt to portray the image as their own.

There are many other techniques for watermark insertion. Watermarks may be implanted into insignificant parts of the document (for instance in images this could be on the boundaries). This is certainly a good solution if the quality of the document is important. The drawback of this technique is that such watermarks can be easily removed. Robust watermarks therefore must be put into significant portion of data. A majority of multimedia watermarking schemes is based on frequency domain manipulation of the multimedia content. Robust watermark aims to modify significant frequency components making the changes difficult to be removed, but at the same time the changes should not visibly degrade the content of the media. Common goals of watermarking are as follows.

- The watermark is invisible and its presence should not be perceptible. This is comparatively easier to achieve due to the redundancy in the multimedia data and the tolerant nature of human vision and hearing.

- The watermarks should be difficult to remove.
- The watermarks should identify the owner of the document.

The second goal, which also implies in itself the third goal, are relatively difficult to assert due to a variety of signal processing tools, such as digital to analog (D/A), analog to digital (A/D), frequency transformation, filtering, equalization, and so forth, are accessible to an attacker. For an image, geometric distortions such as rotation, scaling and cropping, and so forth, may be applied to change the watermarks without noticeable degradation in the image.

In contrast, ***cryptography*** obscures the meaning of a message in such a way that the content becomes unintelligible and meaningless. Cryptography relies on the assumption that the protection is controlled by a secret key whose value is not known to attackers. Cryptography works as long as the attacker does not know the secret key. It is evident that it is not so easy to know some others' secret key. Another advantage of cryptography over steganography is that in case of a successful attack, the security can be easily restored simply by replacing the compromised key with a new one. This is not applicable to steganography as a new hiding scheme must be invented which replaces the compromised one.

Most *digital rights management* solutions make use of encryption in conjunction with digital watermarking techniques so that the information is double protected and ensure better security against illegal copying and distribution. Cryptographic techniques have the capability to provide the standard services (such as confidentiality, authentication, data integrity, and non-repudiation) of network security. In general, encryption techniques are used to prevent illegal copying and distribution, while digital watermarking techniques can be used to establish ownership and discourage copying by allowing grounds for legal actions. Well-designed robust watermarks should be difficult to remove. In case of audio and video files, any intentional or accidental attempts to remove watermarks will result in significant deterioration of the quality of audio and video. A robust watermarking cannot be removed even by repetition of lossy compression (such as JPEG), digital/analog (D/A), and analog/digital (A/D) conversions or any standard signal processing tools such as Fourier transform.

Watermarking the master copy does not prevent production of several identical copies. It also does not control distribution and in case of illegal distribution, trace its source. However, to control illegal copying and prevent the existence of multiple identical copies, it is necessary to have each copy with a unique watermark. So, in this case, existence of multiple copies of the same document indicates a forgery.

There is no ideal watermarking scheme. Consider a scheme based on digital signature. Clearly, attackers are unable to produce a new document with an appropriately modified watermark. However, it is not very difficult to distort the watermark of the document to such a degree that the rightful owner of the document may not be able to claim the ownership.

There are several research works to improve capability of multimedia copyright protection as listed in the references. The future efforts are likely to concentrate on development of new relatively inexpensive hiding techniques (steganography) to be used on massive scale for watermarking of images, video, and sound. Other future work can include extending the protection scheme to support parental rating of the content, variable number of allowed utilizations, time limitations for the content utilization, and similar controlling mechanisms. Interesting ideas of content protection based on biometric data were recently proposed. The methods rely on biometric data of the user and a layered encryption mechanism to achieve confidentiality of the multimedia content (Socek, Sramka, Marques, Oge, & C´ulibrk, 2006).

MULTIMEDIA NETWORKING

Multimedia networks, as the name indicates, are combinations of two basic technologies—networking and multimedia computing. It is a system consisting of connected nodes made to share multimedia data, hardware, and software. Multimedia networking started placing continuous demand on the network infrastructure, and was at odds with packet switching and LAN technologies. One major problem in implementing multimedia applications over TCP/IP has been the problem of delay and jitter. Most applications, which are real-time, cannot tolerate delay with large variations. It makes them unacceptable to the user, because of their unpredictable nature. The major problem arises because of absence of any bandwidth allocation protocols in IP or not setting up any connection to allocate a specific path in UDP. The Internet Engineering Task Force developed Resource Reservation Protocol (RSVP) is an effort to overcome this problem. RSVP over IP allows setup of resource reservations on behalf of an application data stream. RSVP essentially allows a router-based network to simulate circuit-switched network on a best effort basis. When an application requests specific quality of service (QoS) for its data stream, RSVP delivers the request to each router and host states to support the requested level of service. The improved performance and pervasiveness of TCP/IP networks has enabled users to share multimedia information more efficiently across local area and wide area networks, in particular over the Internet.

Lately, asynchronous transfer mode (ATM) has been developed to accommodate the real-time multimedia application issues, specially the delay and jitter problems. By using a 53-byte standard cell size to carry voice, data and video signals the delay problems can be avoided. It can also switch data via hardware, which is more efficient and less expensive. Different traffic types have been defined in ATM each delivering different QoS. One of the traffic types, known as constant bit rate (CBR) is most suitable for multimedia applications. CBR supplies a fixed bandwidth virtual circuit that takes care of delay-sensitive multimedia applications which could be containing real-time video and voice. ATM also provides low latency, high throughput, and scalability which make it a network of choice for supporting new high bandwidth multimedia applications as well as LAN and TCP/IP traffic. ATM speeds are scalable and can exceed 2.5GBPS over fiber.

Cisco, a dominant supplier of data networking hardware for LANs and WANs, offers multimedia software. The aim is to provide all the elements of an end-to-end infrastructure for voice, video, and data traffic across private and public networks.

Corporate Information Superhighway (COINS): Though with the advent of Internet the usage of multimedia applications has reached more people, overcoming the bandwidth limitations still remains a challenge for sometime to come. COINS is a globally connected, fast, efficient, cost-effective, high capacity multimedia network which supports multimedia applications. It is based on a fiber-optic backbone with a capacity of up to 10 gigabits per second to transmit voice, video, data, and images. COINS offer seamless inter-networking, vertical integration, electronic home banking, and electronic commerce. It also provides security, reliability, and is extremely cost-effective. The quest for building an information superhighway, that is, setting up high capacity telecommunications network that would carry vast amounts of digital binary data, is still on.

Multimedia Wireless Networking: For all existing applications, emerging wireless systems will bring a new generation of wireless multimedia applications. The NEC Corporation developed one of the early wireless transmission technologies based on the IEEE1394 high-speed serial bus and capable of 400 megabits, at transmission ranges up to 7 meters through interior wall and up to 12 meters by line-of-sight, which brought multimedia home networking another step closer to reality. The IEEE1394 is well suited to multimedia networking in homes. It has the ability to connect up to 63 devices at a bandwidth of up to 400Mbps and enables a variety of graphics, video, computer and other data to use the network simultaneously. The development of wireless IEEE1394 networking technology now allows for creativity in homes without the hassle of installing new wiring. Some recent research works in multimedia networking area are listed in the references.

The increasing computing power, integrated with multimedia and telecommunication technologies, is bringing into reality our dream of real time, virtually face-to-face interaction with collaborators sitting far away from each other. Multimedia networking promises dramatic improvements in productivity, cost, and user satisfaction, and a generation of new applications.

MULTIMEDIA NETWORKING PROTOCOLS

The use of multimedia is most visible on the Internet. Initially, the use of multimedia was constrained by low network speed, absence of suitable network protocols for transporting multimedia data, and computation limits on both server and client side. Multimedia protocols have been developed to overcome the issues detailed in the previous section. A simplified form of the multimedia protocol stack is shown in Figure 1.

Real-Time Transport Protocol (RTP): It is used on the Internet for transmitting real-time data such as audio and video. The Real-Time Transport Protocol (RTP) does not have a TCP or UDP port to communicate. It runs over UDP via an open port (generally in the range 16384 to 32767) and next higher port (odd) is used for the RTP control protocol (RTCP). UDP can not detect packet loss and restore packet sequence. RTP recover these problems using sequence number and time stamping. It also provides other end-to-end real-time data delivery services that include payload type identification and delivery monitoring. Figure 2 shows a RTP packet format.

The numbers in the parenthesis in Figure 2 indicates the number of bits. The first 96 bits are included in all RTP packets. CSRC identifier is present only when inserted by a mixer. Short description of each of the fixed RTP fields is included in the following.

- *Version (V)* identifies the version of RTP.
- *Padding (P)* if set indicates that the packet contains one or more additional padding octets at the end which are not part of the payload.
- *Extension (X)* if set indicates that the fixed header is followed by exactly one header extension.
- *CSRC count (CC)* contains the number of CSRC identifiers that follow the fixed header.
- *Marker (M)* is intended to allow significant events such as frame boundaries to be marked in the packet stream.
- *Payload Type (PT)* identifies the format of the RTP payload and determines its interpretation by the application.
- *Sequence Number* increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence.
- *Timestamp* reflects the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations
- *SSRC* field identifies the synchronization source.

Figure 1. Internet multimedia protocol stack

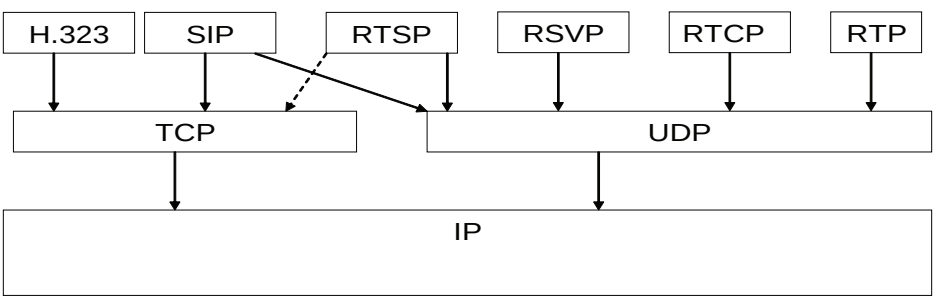
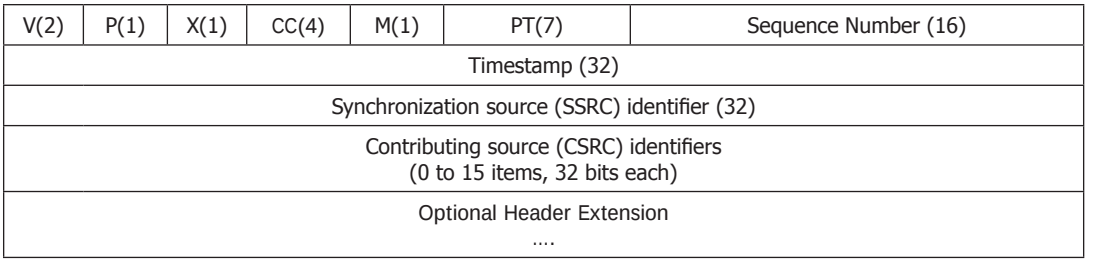


Figure 2. RTP packet format



Detailed RTP format can be found in RFC3550 (http://tools.ietf.org/html/rfc3550) and payload can be found in RFC 2038 (http://faqs.org/rfcs/rfc2038.html).

Real-Time Control Protocol (RTCP): It is an Internet protocol that works in conjunction with RTP to monitor the quality of service and to convey information about the participants in an on-going session. The RTP control protocol (RTCP) is based on the periodic transmission of control packets to all participants in the session, using the same distribution mechanism as the data packets. Feedback can be used to control performance. Sender may modify its transmissions based on feedback. Each RTCP packet contains sender and/or receiver reports. Statistics include number of packets sent, number of packets lost, inter-arrival jitter, and so forth. For more details see http://www.freesoft.org/CIE/RFC/1889/14.htm.

Resource Reservation Protocol (RSVP): A host uses RSVP to request a specific quality of service (QoS) from the network, on behalf of an application data stream. RSVP carries the request through the network, visiting each node the network uses to carry the stream. At each node, RSVP attempts to make a resource reservation for the stream. RSVP does not perform its own routing; instead it uses underlying routing protocols to determine where it should carry reservation requests. There are seven messages used in RSVP: Path, Resv, Path Teardown, Resv Teardown, Path Error, Resv Error, and Confirmation. The RSVP protocol is described in RFC 2205, updated by RFCs 2750, 3936, 4495 (http://tools.ietf.org/html/rfc2205).

Real-Time Streaming Protocol (RTSP): It is a protocol for use in streaming media systems which allows a client to remotely control a streaming media server, issuing VCR-like commands (i.e., pause/resume, repositioning of playback, fast forward, and rewind), and allowing time-based access to files on a server.

Most RTSP servers use the RTP as the transport protocol for the actual audio/video data. However, a proprietary transport protocol, known as *Real Data Transport (RDT)* developed by RealNetworks, is used as the transport protocol for RTSP server from RealNetworks. The RSVP protocol is described in RFC 2326 (http://tools.ietf.org/html/rfc2326).

The H.323, SIP (for call control and signaling), and RTP, RTCP, RTSP (for audio/video) are the protocols and standards used for Internet Telephony. SIP is a signaling protocol that initiates, manages, and terminates multimedia sessions. H.323 supports H.245 over UDP/TCP and Q.931 over UDP/TCP and RAS over UDP.

MULTIMEDIA NETWORKING RESEARCH

As emerging multimedia technologies are providing higher performance available at competitive prices, they are also enabling and proliferating multimedia solutions in a spectrum of commercial and laboratory projects. Advances in computer and networking technologies have fueled the rapid growth of research and development in multimedia computing and high-speed networking. Some common research areas include:

- Development and management of real-time distributed multimedia applications;
- Audio/video applications and streaming issues;
- Protocols and technologies for building and transporting multimedia applications on the Internet;
- QOS frameworks and implementation;
- Collaborative applications;
- Multimedia synchronization in distributed environment;
- Multicasting technology and applications;
- Use of mobile multimedia over wireless network; and
- Dynamic and efficient usage of resources.

Real-time distributed multimedia environments, characterized by timing constraints and end-to-end quality of service (QoS) requirements is a challenge for the efficient management mechanisms to respond to transient changes in the load or the availability of the resources (Andreadis & Zambon, 2007).

QOS Frameworks and Implementation: In multimedia applications, media data such as audio and video are transmitted from server to clients via network according to some transmission schedules. Different from the conventional data streams, end-to-end quality-of-service (QoS) is necessary for media transmission to provide jitter-free playback. Several researchers, while dealing with the protocol and technology aspects, also have focused on the QOS frameworks and implementation issues.

Because of a wide range of application areas, the delivery of high quality video content to customers is a driving force for the evolution of the Internet (Datta, Li, & Wang, 2005; Yang, Li, & Zhuang, 2002). There are video retrieval applications with flexible user interfaces based on HTTP browser for content querying and browsing, support for both unicast and multicast addressing and a user-oriented control of QoS of video streaming in Integrated Services IP networks. Information systems' modelling methods required for the prediction of system performance and an influence of different control mechanisms on quality of service perceived by end users are related areas of research (Nang & Park, 2007; Rowe & Jain, 2005; Sebe & Tian, 2007).

The emergence of high-speed networked multimedia systems opens up the possibility that a much more diverse collection of continuous media (CM) applications could be handled in real time. Admission control in CM servers or video-on-demand systems restricts the number of applications. It is necessary to develop more intelligent mechanisms for efficient admission control, negotiation, resource allocation, and resource scheduling with an aim to optimize the total system utilization.

The traffic generated by multimedia applications presents a great amount of burstiness, which can hardly be described by a static set of traffic parameters. For dynamic and efficient usage of the resources the traffic specification should reflect the real traffic demand and at the same time optimize the resources requested. There are models, as discussed in the previous section, for dynamically renegotiating the traffic specification (RVBR) and integrating with the traffic reservation mechanism RSVP demonstrating through an example of application that is able to accommodate its traffic to manage QoS dynamically. (Syed, 2002).

SECURING MULTIMEDIA NETWORK

The security of multimedia data transmitted over a network is important for multimedia commerce. For example, in video on demand and video conferencing applications, it is desirable that only those who have paid for the services can view their video or movies. There is research suggesting different schemes to secure multimedia networks via user identities. Some scheme is composed of multimedia key distribution, multimedia encryption, and multimedia authentication to ensure that the multimedia data can be securely and efficiently distributed to multimedia users on the basis of their identities.

Authentication control mechanisms can be used to secure distributed multimedia applications. However, it is not enough to secure multimedia data broadcast on wireless, satellite or Mbone networks. Multimedia data is still needed to be encrypted during transmission. Encryption algorithms can be divided into two basic classes—*secret-key and public-key encryption* algorithms. They have distinct characteristics and are used in different ways to provide security services.

Secret-key encryption algorithms have been in use in commercial networks since the early 1970s. The U.S. Data Encryption Standard (DES) is the first secret-key encryption algorithm which has had its full specification published as a public standard. It was developed at IBM in 1976. It encrypts 64-bit blocks of data with a 56-bit key. Considering the disagreement over whether a 56-bit key is sufficiently strong, a number of secret-key encryption algorithms have been proposed to replace DES in recent years. The International Data Encryption Algorithm (IDEA), developed by Xuejia Lai and Jame Massey in 1990, is one of them. IDEA encrypts 64-bit blocks of data with 128-bit key. Another class of secret-key encryption algorithm is the stream cipher which uses a short key to generate the key-stream to encrypt a digital data stream one bit at a time.

Public-key encryption algorithms are based on mathematical functions rather than on substitution and permutation. It is asymmetric involving the use of two separate keys, in contrast to symmetric secret-key algorithm which uses only one key. The use of two keys has profound consequences in the areas of confidentiality, key distribution, and authentication. State-of-the-art public-key encryption algorithms with high security transmission performance require high processing resources when applied to high bit-rates and result not suitable for modern multimedia communications. Although existing secret-key encryption algorithms, such as DES, operate much faster than public-key algorithms, they are very complicated and involve large computations. A software DES implementation is not fast enough to process the vast amount of data generated by multimedia applications and a hardware DES implementation (a set-top box) adds extra costs both to broadcasters and to receivers.

Multimedia data security is challenging that come from two facts. First, multimedia data size is usually very large. Second, multimedia data needs to be processed in real time. Encryption algorithms with high security will put a great burden on storage space requirement and increase latency. For most multimedia applications, the information rate is very high, but the information value is very low. For some commercial applications, such as pay-per-view, very expensive attacks on the encrypted multimedia data are not interesting to the attacker because of their low information value. Breaking such encryption code is much more expensive than buying the services.

Some encryption algorithms with high security, such as DES, IDEA, and AES, when applied to high throughput multimedia data, will put great burden on storage space demand and increase latency. Thus, they may not be suitable for multimedia communications. The study related to fast video encryption can be found in many of the referenced articles, but these algorithms are only applied to video data. In view of this fact, efficient encryption algorithms for any multimedia data with appropriate security are needed. Digital signatures can be used for a multimedia user to verify multimedia data integrity which is endorsed by a multimedia server. Usually, digital signatures are produced with Digital Signature

Standard. In order to verify a DSS digital signature, the public key certificate of the signer should be submitted to the verifier for verification. Encryption algorithms and recent research has been covered in several works listed in the reference section.

Fast encryption algorithm for multimedia data, FEA-M, has been proposed by some researchers to overcome these issues. It encrypts 512 bytes of data with 512 bytes of key at a time. Cryptanalysis of this algorithm has shown that FEA-M satisfies basic confusion and diffusion design criteria. The structure of FEA-M facilitates both hardware and software implementations. Computation complexity comparison has shown that FEA-M is much faster encryption algorithm than others. It needs only about 1.5 XOR operations to encrypt one bit plaintext. FEA-M can be used to secure many multimedia applications, such as digital video and audio transmissions (Yi, Tan, Siew, & Syed, 2002).

MULTIMEDIA APPLICATIONS

We are witnessing an explosive growth in the use of multimedia in diverse application areas such as corporate presentations, communication, collaborative work, information delivery, sales and merchandising, entertainment, training and education both in distance and face to face mode, electronic commerce, simulations, digital publications, museum exhibits, games and multimedia interactive kiosks, and so forth. The application of multimedia has expanded to home libraries, magazines and newspapers, and classroom teaching. It is multimedia that is making it possible for us to remain in our homes and shop for products and services of our choices at prices of our liking. Businesses can also interact with their consumers only through their Web presence.

Some of the advanced applications of multimedia, which were otherwise thought of just as ideas, are a reality today. Some examples include videoconferencing, video phone, iPod, applications on personal digital assistants, and mobile telephones. One concept that is making fast progress is the interactive television, though it is not being used as widely because of the costs involved in moving data to and fro between the user and the service provider. If the demand for this service increases then the costs would certainly come down considerably. It is not far in future when there will be multimedia homes with virtual conference going on in one room, kids doing their homework from online databases or kids playing virtual reality games. Smart devices like portable newspapers and interactive televisions will be common for all.

In short, it will be difficult to name any application in our daily life that does not involve multimedia. In this section few of the commonly used applications are discussed.

In-Vehicle Multimedia Systems: With the growth of multimedia it is now a reality to envision an entertainment network in any automobile. Innovative A/V features are currently available on rear seat and passenger side networks, which is a driving force to bring multimedia throughout the vehicle. In-vehicle networks allow for connectivity between a vehicle and consumer electronic device. The high bandwidth, quality, and reliability requirements of multimedia have caused a demand for standardization. IDB-1394 is an international data networking standard for transmitting video, audio, and other multimedia data over an in-vehicle network that provides the bandwidth required for these applications. This is a standard suitable to enable rear-seat entertainment systems that can play DVDs, show DTV programming, and offer access to the car navigation system. The IDB-1394 standard provides a minimum bit rate of 400 megabits/second that is higher than other alternative standard such as Ethernet or MOST (which stands for Media Oriented Systems Transport). The future plan is to move IDB-1394 to 800 megabits/second, make it scalable and versatile. It will also be compatible to preserve current 1394-equipped product investment as speed and bandwidth increase.

MOST is the emerging fiber-optic network standard optimized for in-vehicle multimedia and entertainment devices. The network was founded in 1998 by many of the leading automotive manufacturers and vendors developing telematics solutions for the automobile industry, including Audi, BMW, DaimlerChrysler, Harman/Becker, Motorola, Oasis Silicon Systems, Johnson Controls, and Delphi-Delco. MOST is rapidly becoming a de facto auto industry standard for building multimedia networking products.

Yazaki North America Inc., as one of the leading provider of vehicle power and data network (PDN) solutions for the automotive industry, decided in mid-2002 to use Microsoft Windows CE in the development of in-vehicle multimedia systems. By incorporating Windows CE for Automotive, the drivers can access the in-vehicle network to control an array of potential applications, such as voice recognition and voice controls, navigation, digital video and audio, satellite TV, and video games. The MOST audio/video gateway converts data from both analog and digital sources to MOST digital data streams, providing cost-effective connectivity for portable DVD, CD, MD, and MP3 players, as well as other audio and video devices. Fully scalable, the MOST audio/video gateway also provides connectivity to CAN- and J1850-enabled devices such as radios, amplifiers, and CD changers, allowing designers to use legacy systems and components within the MOST multimedia network.

The conceptual model of Yazaki power and data networks (PDNs) consists of the following four-layer network infrastructure based on broad functionality, data-transmission speed, and corresponding data protocols. A given layer may be made up of two or more sub-networks.

- **Infotainment Layer:** It contains advanced audio and video devices, wireless phones, voice-activated systems, instrumentation sensing, and other systems requiring SAE Class D data. Transfer speed is greater than 1Mbps. It is based on the MOST data protocol.
- **Body Layer:** It provides control of lighting, HVAC, door and seat functions as well as speed control, and instrumentation. These functions are supported by low-speed, SAE Class A networks operating at lower than 10 kb/s, and SAE Class B networks running at 10-125 kb/s.
- **Safety and Mobility Layer:** It includes airbags and engine control modules plus steering, transmission, and x-by-wire systems. These systems are supported by SAE Class C and Class D networks, which provide real-time control, with operating speeds from 125 kb/s to 1 mb/s and greater.
- **Energy Layer:** It includes power generation, storage, and distribution elements (typically an alternator, battery, and under hood power distribution component).

Recently (November 2007) Fujitsu Microelectronics Asia Pte Ltd. has developed controllers, which support the IDB-1394 standard for in-vehicle multimedia networks. SmartCODEC, developed by Fujitsu Laboratories Ltd., can transmit high-resolution video without perceptible latencies, and it is possible to realize high-quality, cost-effective rear-seat entertainment. It allows transmitting video from DVD, DTV, and car navigation sources at high resolution without perceptible lag. Fujitsu plans also to expand its range of IDB-1394-compliant offerings, in view of the growth types of in-vehicle content, and to accommodate multiple data transmissions from peripheral car security cameras, and further system cost reductions (http://news.soft32.com/fujitsu-to-launch-new-chips-for-in-vehicle-multimedia-functions_5509.html?pid=5509&rate=5).

Advances in mobile communication technologies make it possible to receive multimedia content from different broadcasters. In the case of public transportation, such as in buses, trains, and airplanes, the in-vehicle multimedia content is viewed by on-board passengers. Accordingly, adaptive in-vehicle multimedia recommendation systems are useful in such situations to serve for group users. Research are being done to develop adaptive vehicular multimedia systems to provide personalized multimedia content for group users by taking care of the majority's preferences (Zhiwen, Xingshe, & Daqing, 2005).

Videoconferencing Systems: Videoconferencing is conducting a conference between two or more participants at different sites by using computer networks to transmit audio and video data. Videoconferencing systems are in place for quite some time. With the development of technology for transmission of streaming audio and video, PC-based videoconferencing, or satellite TV-based broadcast systems is becoming more affordable to general users. Until recently these were costly products accessible only to large organizations or by renting. Videoconferencing is still considered a poor alternative to face-to-face meetings. Videoconferencing systems are often used for group-to-group meetings where spatial distortions are exacerbated. In the business setting, where these systems are most prevalent, the misuse of videoconferencing systems can have detrimental results, especially in high-stakes communications. Its effects on the group dynamic are not well understood. Lots of research is being done to improve their performances.

A similar class of application that promises on-demand access to multimedia information is radio and broadcast news. Works in this areas show how the synergy of speech, language, and image processing has enabled a new class of information on demand news systems. It has the ability to automatically process broadcast video 24/7 and serve the general public in individually tailored personal casts. A medical computer supported cooperative work (CSCW) tool for medical teleconsultation supported by an open tele-cooperation architecture with modern high power workstations was implemented using distributed computing system (Holmberg, Wünsche, & Tempero, 2006).

Multimedia Office/Business: Images, animation, and audio can be added to even the basic office applications like a word processing package or a spreadsheet to emphasize important points in the documents and make them powerful tools. One of the most common uses is in multimedia presentations which are a great way to introduce new concepts or explain a new technology and reduce the design and training time of multimedia in companies. Major applications include may include: executive information systems, electronic publishing, remote consulting systems, multimedia mail, advertising, collaborative work (Jiang, Wang, & Benbasat, 2005; Yu, 2004). There is no doubt that multimedia has changed the way we do business. However, the impact of few applications, such as tele-conferencing, real-time, online, video-link, and so forth, was not as profound as was expected. Videoconference did not replace all business meetings. People would much rather do business face-to-face to feel comfortable and natural than over a video link, even at the cost of few hour's flight.

Multimedia Entertainment: The field of *entertainment uses multimedia* extensively. One of the earliest applications of multimedia was for games. Multimedia made possible innovative and interactive games that greatly enhanced the learning experience. Games become alive with sounds and animated graphics. Examples are game-show formats, hangman games, drag and drop exercises, and puzzles where students complete against the computer, another student, or the student's past performance (Chooprayoon & Fung, 2007; Magerkurth, Engelke, & Grollman, 2006; Magerkurth, Engelke, & Memisoglu, 2004).

The entertainment industry has greatly benefited from multimedia technology mainly because of their high sound and picture quality. In many occasions music is entirely created by computers and synthesis. Even most Hollywood movies now use some sort of digital imagery or editing within it. 3-D animated movies, such as Jimmy Neutron, Toy Story, and Ice Age, were created using the 3-D multimedia development programs.

Multimedia Simulation: Simulation equipped with multimedia imitates real life, situations, tasks, or procedures. Flight simulators or driving simulators may be used for training pilots or drivers without having an airplane or a car. In most cases simulation is used to train or practice decision making without having a very expensive or dangerous equipment or environment. For example, this might be used to teach manipulation of dangerous chemicals or the transporting of missiles or bombs. This type of instruction can be time-consuming and therefore costly to design and develop. Interactive simulations are the highest level of MM enhancements (Rocceti & Syed, 2003).

Virtual reality is a truly absorbing multimedia application. It is an artificial environment created with computer hardware and software. It is presented to the user in such a way that it appears and feels real. In virtual reality, the computer controls three of the five senses. Virtual reality systems require extremely expensive hardware and software and are confined mostly to research laboratories (Naef, Staadt, & Gross, 2004)

E-Commerce: The 21st century business environment is dynamic and competitive. The technology world forces on business operations all around the world to change traditional methods of business to Internet-based electronic commerce (e-commerce). E-commerce offers the most effective business opportunities in the marketplace. Electronic commerce is essential for the survival of companies in the virtual distribution marketplace.

Multimedia in Education: Delivery of content in education has gone through tremendous improvement taking advantage of multimedia. Research on relationship between cognition, learning and education through answering questions like "how does a person approach the learning of a new concept" or "how can one remember things more effectively" established positive impacts of multimedia on each step of learning cycle that involves attention, rehearsal, encoding, retrieval, attention. In the early days text-based computer assisted instruction (CAI) systems were used for delivering lectures. With inclusion of multimedia significant improvements were experienced in grabbing the attention of the learner. Lately interactive multimedia with non-linear presentations provided substantial interaction with the user. The user establishes the concept in mind more firmly by receiving feedback of the activities performed for different levels of difficulty selected by the user or the system based on the evaluated skills. Application of multimedia is varied and may be accommodated for different learning styles (e.g., association vs. experimentation; auditory vs. visual) at different level with different expectations and outcomes. For example, multimedia lessons may be developed for children aimed at visual memory development, which is a very important step in the development of children in the early ages. Different issues related to multimedia in education and recent research has been covered in several works listed in the reference section.

Multimedia Network for Healthcare: One leading healthcare provider, the Northwestern Memorial Hospital, deployed a next generation 3Com multimedia network consisting of industry-leading enterprise switches and supporting systems to access patient-critical medical applications and data. The aim was to improve medical, scheduling, and billing processes to bolster patient care and satisfaction. The system was designed to provide the staff with up-to-the-minute patient records and diagnoses and the most sophisticated medical services available. The network would provide caregivers with a fast and reliable flow of patient information for timely and effective healthcare and also support the newest multimedia teaching technologies. The network system may also be incorporated into the enhancing and ongoing training of medical students with multimedia applications. It also has important features that the students are able to view live surgeries and other operating room procedures for their desktops (Cucchiara, 2005; Ebadollahi, Coden, Tanenblatt, Chang, Mahmood, & Amir, 2006; Halle & Kikinis, 2004; Kulkarni & Öztürk, 2007).

Voice over Internet Protocol (VoIP) over digital subscriber line (DSL) offers the ability to provide both voice and data over a single copper loop, cost effectively and with flexibility and functionality. DSL's dedicated lines provide increased security and guaranteed bandwidth that is comparable to that offered by T-1 and frame relay circuits but at a lower cost.

Internet Telephony: Internet telephony is the technology that makes it possible to have a telephone conversation over the Internet. This technology also includes voicemail online and conference calls without being tied to any area code. Different Internet phone services are: PC-to-PC, PC-to-Phone, and Phone-to-Phone. Until recently, Internet telephony was characterized with poor voice quality and long

time delays in transmission. With elimination of these problems Internet telephony's voice quality has become competitive to its rival PSTN. Internet phone services have grown from technical novelty to a competitive threat for traditional circuit-switched telecommunications. It offers one common service combining voice, video, and data traffic by adopting IP as a common protocol and merging different network structures in one comprehensive medium. While most Internet applications are seen as another or additional source of revenue to the Interexchange Carriers (IXC's), Internet telephony, or the transport of voice over the Internet, it might pose a potential threat from the view point that the circuit-switched technology will be obsolete and be considered for an updated technology. Traditional carriers and telcos are beginning to feel similar pressure predicted by Forrester Research that by the year 2004, U.S. telephone companies alone would have lost some $3 billion to Internet telephony. From the users' point of view, it might mean an improved service at a steeply discounted rate, especially for long-distance calls.

With the availability of modern development tools, it will not be a big mistake to make assumption that today the multimedia end product is only limited imagination of the developer and requirement specification of the customer.

PEER-TO-PEER (P2P) MULTIMEDIA NETWORK APPLICATIONS

In both academia and industry, peer-to-peer (P2P) applications have attracted great attention. In a peer-to-peer (P2P) overlay network, a large number and heterogeneous types of peer processes are interconnected in networks (having wide varieties of computing and network resources) with an aim to exchange multimedia contents, such as movies, music, pictures, and animations, and so forth, in a reliable and real-time manner. It is different from client-server based systems in a way that the peers bring with them server capacity. Multimedia streaming is a key technology to realize multimedia applications in networks. The P2P streaming applications, such as PPLive, UUSee, are on the rise. These enable the P2P file sharing/streaming application inexpensive to build and excellent in scalability. Following are some of the successful and well-known P2P file sharing applications.

- BitTorrent
- Gnutella
- Kazaa
- Napster
- PPLive
- Skype

Some statistics indicate that P2P traffic accounts for almost 70% of Internet traffic.

MULTIMEDIA DATA MINING/MULTIMEDIA RETRIEVAL

Multimedia and data mining are two very interdisciplinary and multidisciplinary areas that are experiencing rapid developments in recent years. While multimedia is facing the challenge in handling data due to the enormous growth in content and data mining has become a popular way of discovering new knowledge from large data sets, accordingly multimedia data mining emerged as a discipline that brings together database systems, artificial intelligence, and multimedia processing. Multimedia retrieval/multimedia mining makes it possible to retrieve desired information, analyze characteristics of an information set,

and discover knowledge hidden in vast amounts of multimedia information. It is, therefore, is a technology necessary for the effective use of multimedia information. Analysis and/or mining of knowledge are very important for marketing, product design, science domains, and various other areas. Museum artifact retrieval systems and video retrieval systems in the education field can be very effective when equipped with data mining capabilities.

ISSUES, TRENDS, AND EFFECTS OF MULTIMEDIA

With the increased availability and ease of access to electronic multimedia information, new challenges are arising in all areas such as data management, retrieval, synchronization, and transportation of large volumes of media generated data. Also we need to address a number of technology, management, and design issues.

Managing Multimedia Projects: Organizing and managing multimedia IT project remains a challenge. Creativity is an important key word in developing multimedia IT. It is very difficult to measure the level of creativity in each of the players of a project, namely the employees who are involved in creating and delivering the project within the given dead line and the clients with varying expectation of the creativity. For a desired good output the employers need to ensure a creative environment so that the employees can think creative and work at their own pace. The fact remains that "some employees can turn on creativity like a light switch and others need creativity to come to them".

Interoperability is also a very important issue. The applications need to be coded in one standardized coding format, making them accessible to a whole range of end-systems supporting real-time audio/video stream and synchronization. There also need to be a set of protocols defined to implement transfer of all messages and dataflow in the system. Gateways will have to be specified to integrate various digital services and support the corresponding formats.

In multimedia applications, digital imagery and video has expanded its horizon in many directions, resulting in an explosion in the volume of image, audio, and video data required to be organized and retrieved. Accordingly, we are witnessing a great increase in research on multimedia retrieval, which has paved the way for new techniques and fascinating applications (Datta, Li, & Wang, 2005; Nang & Park, 2007; Rowe & Jain, 2005; Sebe & Tian, 2007; Yang, Li, & Zhuang, 2002).

Managing Multimedia Resources: Because of large file sizes of audio and video contents, communication over the Internet involving multimedia data consumes lots of network bandwidth, occupies large storage space, and requires high processing time both on the client and on the server side. With large multimedia files the maximum serving capacity of a server decreases exponentially, which is true even in case of an infinitely fast server. It is said, "The backbone of the Internet may have been designed to withstand nuclear assault, but it will never survive the onslaught of multimedia on the Web". The large multimedia file sizes not only consume more bandwidth during transmission, they occupy connections for long time, which causes delay to other real-time data transfers and also decrease the total number of connections available at any one time. In the long run, these connections make it easier for a server to run out of TCP/IP kernel resources (http://www.webdeveloper.com/multimedia/multimedia_dark_side.html).

Trends in Mobile Multimedia: During the last decade the networking speed has grown from 10 megabit to gigabit range, but the explosive growth in the number of the Internet users has caused the amount of traffic to grow several times more. At the same time due to the significant advances in the VLSI technology, there is an increasing demand for portable multimedia appliances capable of handling advanced algorithms required in all forms of communication. Over the years, we have witnessed a steady

move from stand-alone (or desktop) multimedia to deeply distributed multimedia systems. Wireless computing is taking the world by storm to serve our information need. So, there is a need for efficient mapping of the requirements of multimedia systems onto mobile networking environments. The success criteria for a mobile application design is to find the best mapping onto the architectural resources, while satisfying an imposed set of design constraints (e.g., minimum power dissipation, maximum performance) and specified QoS metrics (e.g., end-to-end latency, jitter, loss rate) which directly impact the media quality. Applications are evolving in order to support today's mobilized workplace and lifestyle. But multimedia systems yet need to become more adaptive or scalable with respect to the fluctuating network environment (Cai, Shen, & Mark, 2006; Farnham, Sooriyabandara, & Efthymiou, 2007; Fernando, 2006; Koutsakis, 2007; Pham & Wong, 2004; Thwaites, 2006; Verkasalo, 2006).

Trend in Multimedia Research: Multimedia research will continue to address the crucial problems that are evolving every day as attempts are being made to bring the dream applications into reality. Some recent research is included in the reference list. Research works in multimedia has been extended from software development to concept of probabilistic design for multimedia embedded systems, which systematically would incorporate performance requirements of multimedia application, uncertainties in execution time, and tolerance for reasonable execution failures (Hua, Qu, & Bhattacharyya, 2007). Because of the dominant computing workload of multimedia applications in computer systems and in wireless-based devices and due to their repetitive computing and memory intensive nature, Lanuzza, Margala, and Corsonello (2005) suggested to take effective advantage from processor-in-memory (PIM) technology and proposed a new low-power PIM-based 32-bit reconfigurable datapath optimized for multimedia applications.

Social Impact: One other important aspect is the social implications that multimedia-based applications may have. The positive aspects are more obvious than the negative ones. With so many facilities readily available at homes, offices, and shops, there are several questions remain yet to be answered such as, how the reduction in human contact will affect the social behavior—will the offices, the playgrounds, or the malls have a deserted look? Because, the people prefer to work from home or they would spend more time on games and virtual reality avoiding outdoor exercises, breathing fresh air or they prefer ordering through interactive television while watching a movie. There is always a balance which can be found to benefit from a developing technology like multimedia and it sure will make information and communication easy to access and use.

Human-Centered Multimedia—Cultural Impact: The wide acceptance of multimedia may be credited to the human-centered characters inherent in media combinations. However, the cultural part of the human-centered approach did not get much consideration in design of multimedia development tools and applications. We often forget that the development of multimedia interfaces and communication are language and culture-specific. The users feel much comfortable when they can act to the cultural context to which they belong (Arts, 2004; Dimitrova, 2004; Rowe & Jain, 2005). Examples of how culture affects content production and automatic analysis techniques may be found at http://www-nlpir.nist.gov/projects/trecvid/ (The TREC video retrieval evaluation, 2005). Recent trends are to align technical development with social and cultural development involving new models that embraces multiculturalism and recognizes the human impact.

CONCLUSION

In the span of the last two decades, we have witnessed dramatic increases in the use of digital technologies for information storage, processing, and delivery. The potential of multimedia has demonstrated

to be one of the most fascinating and fastest growing areas in the field of information technology. The current technological impacts of multimedia are much larger then most people think it is. Until even recently it was not unusual to be frustrated with a multimedia project running on a slower computer because it didn't run at a desired speed with desired quality. Multimedia consumes a lot of computer power to run effectively and influences greatly in the decision of people when purchasing new computers. Now many people can afford to get the latest and fastest computer for video editing of their home movies. Businesses can now understand their markets more in depth than ever before enabling them to improve quality of production, products, services, and consumer satisfaction.

Multimedia is playing and will continue to play a very important role in today's visual society. Interactive multimedia will grow in maturity and popularity on the gigabit networks. With technologies continuing to push toward the seamless integration of multimedia and the Internet, computer speed gaining a super-speed, promise of gigabit home networks looming on the horizon and improved compression techniques emerging at a fast pace, we can confidently conclude that multimedia productions are yet to show their full potential—the multimedia of the future.

REFERENCES

Aiello, W., Bellovin, S. M., Blaze, M., Canetti, R., Ioannidis, J., Keromytis, A. D., & Reingold, O. (2004). Just fast keying: Key agreement in a hostile internet. *ACM Transactions on Information and System Security, 7*(2), 130.

Anderer, C., Neff, J. M., & Hyde, P. (2007). Multimedia magic: Moving beyond text. *Proceedings of the 35th Annual ACM SIGUCCS Conference on User Services SIGUCCS '07.*

Andreadis, A., & Zambon, R. (2007). QoS scheduling and multimedia: Qos enhancement for multimedia traffics with dynamic txoplimit in IEEE 802.11e. *Proceedings of the 3rd ACM Workshop on QoS and Security for Wireless and Mobile Networks Q2SWinet '07.*

Antonietti, A., & Giorgetti, M. (2006). Teachers' beliefs about learning from multimedia. *Computers in Human Behavior, 22*(2), 267-282.

Arts, E. (2004). Ambient intelligence: A multimedia perspective. *IEEE Multimedia, 11*(1), 12-19.

Becvar, L. A. (2007). *Social impacts of a video blogging system for clinical instruction.* CHI 2007, ACM Student Research Competition, San Jose, CA, USA.

Bergman, D. S., Youssef, B. B., Bizzocchi, J., & Bowes, J. (2006). Interpolation techniques for the artificial construction of video slow motion in the postproduction process. *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology ACE'06.*

Bolettieri, P., Falchi, F., Gennaro, C., & Fausto, R. (2007). *Automatic metadata extraction and indexing for reusing e-learning multimedia objects.* Workshop on Multimedia Information Retrieval on the Many Faces of Multimedia Semantics MS '07.

Brosnan, E., Fitzpatrick, C., Sharry, J., & Boyle, R. (2006). An evaluation of the integrated use of a multimedia storytelling system within a psychotherapy intervention for adolescents. *CHI '06 Extended Abstracts on Human Factors in Computing Systems CHI.*

Brotherton, J. A., & Abowd, G. D. (2004). Lessons learned from eclass assessing automated capture and access in the classroom. *ACM Transactions on Computer-Human Interaction (TOCHI), 11*(2), 121-155.

Buttussi, F., Chittaro, L., & Coppo, M. (2007). Using Web3D technologies for visualization and search of signs in an international sign language dictionary. *Proceedings of the Twelfth International Conference on 3D Web Technology Web3D '07.*

Cai, L. X., Cai, L., Shen, X., & Mark, J. W. (2006). Multimedia capacity of UWB networks supporting multimedia services. *Proceedings of the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks QShine '06.*

Chooprayoon, V., & Fung, C. C. (2007). Thai small and medium enterprises engagement in the emerging digital content and game industry. *Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts DIMEA '07.*

Christine, L. L. (2007). *Collaboration and dialogue: Video as an agent in extending and enriching learning and writing.* ACM SIGGRAPH 2007 Educators Program SIGGRAPH '07.

Churchill, E. F., Nelson, L. L. D., Helfman, J., & Paul, M. (2004). Interactive systems in public places: Sharing multimedia content with interactive public displays: A case study. *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques DIS.*

Cucchiara, R. (2005). Image and video processing for multimedia surveillance systems: Multimedia surveillance systems. *Proceedings of the Third ACM International Workshop on Video Surveillance & Sensor Networks VSSN '05.*

Datta, R., Li, J., & Wang, J. Z. (2005). Multimedia information retrieval: Challenges and real-world applications: Content-based image retrieval: approaches and trends of the new age. *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval MIR'05.*

Davis, H. C., & White, S. (2005). A research-led curriculum in multimedia: Learning about convergence, ACM SIGCSE Bulletin. *Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education ITiCSE '05, 37*(3).

Dimitrova, N. (2004). Context and memory in multimedia context analysis. *IEEE Multimedia, 11*(3), 7-11.

DiPaola, S., & Akai, C. (2006). *Designing an adaptive multimedia interactive to support shared learning experiences (panel)*. ACM SIGGRAPH 2006 Educators Program SIGGRAPH '06.

Dittman, J., Wohlmacher, P., & Nahrstedt, K. (2001). Using cryptographic and watermarking algorithms. *IEEE Multimedia, 8*(3), 54-65.

Ebadollahi, S., Coden, A. R., Tanenblatt, M. A., Chang, S. F., Mahmood, T. S., & Amir, A. (2006). Multimedia signal processing and systems in healthcare and life science: Concept-based electronic health records: Opportunities and challenges. *Proceedings of the 14th Annual ACM International Conference on Multimedia MULTIMEDIA '06.*

Farnham, T., Sooriyabandara, M., & Efthymiou, C. (2007). Enhancing multimedia streaming over existing wireless LAN technology using the unified link layer API. *International Journal of Network Management, 17*(5).

Favier, F., Fynn, J., & Misson, M. (2007). Introducing wireless networking technologies as a teaching tool for foreign languages: A multimedia laboratory experiment. *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education Emme '07.*

Fernando, X. (2006). Radio over fiber in multimedia access networks. *Proceedings of the 1st International Conference on Access Networks AccessNets '06.*

Feuvre, J. L., Concolato, C., & Moissinac, J. (2007*).* Open source software competition: GPAC: Open source multimedia framework. *Proceedings of the 15th International Conference on Multimedia MULTIMEDIA '07.*

FIPS PUB 197. (2001). *Advanced encryption standard.* Federal Information Processing Standards Publications, U. S. Department of Commerce/N.I.S.T., National Technical Information Service.

Franco, J. F., Cruz, S. R., & Lopes, R. D. (2006). *Computer graphics, interactive technologies and collaborative learning synergy supporting individuals' skills development (panel*). ACM SIGGRAPH 2006 Educators Program SIGGRAPH '06.

Friedland, G., Hürst, W., & Knipping, L. (2007). Educational multimedia systems: The past, the present, and a glimpse into the future. *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education Emme '07.*

Furht, B., Muharemagic, E. A., & Socek, D. (2005). Multimedia security: Encryption and watermarking. *Multimedia Systems and Applications, 28*.

Halle, M. W., & Kikinis, R. (2004). Multimedia in life and health sciences. Flexible frameworks for medical multimedia. *Proceedings of the 12th Annual ACM International Conference on Multimedia MULTIMEDIA '04.*

Holmberg, N., Wünsche, B., & Tempero, E. (2006). A framework for interactive Web-based visualization. *Proceedings of the 7th Australasian User Interface Conference, Volume 50, AUIC '06*. Australian Computer Society, Inc.

Hua, S., Qu, G., & Bhattacharyya, S. S. (2007). Probabilistic design of multimedia embedded systems. *ACM Transactions on Embedded Computing Systems (TECS), 6*(3).

Jaimes, A., Sebe, N., & Gatica-Perez, D. (2006). Human-centered multimedia: Human-centered computing: A multimedia perspective. *Proceedings of the 14th Annual ACM International Conference on Multimedia MULTIMEDIA '06.*

Jain, R. (2005). Networking and mobile computing: Improving quality of service for streaming multimedia applications in ubiquitous mobile environment. *Proceedings of the 43rd Annual Southeast Regional Conference, Volume 2, ACM-SE 43.*

Jäkälä, M., & Pekkola, S. (2007). From technology engineering to social engineering: 15 years of research on virtual worlds. *ACM SIGMIS Database, 38*(4).

Jiang, Z., Wang, W., & Benbasat, I. (2005). Multimedia-based interactive advising technology for online consumer decision support. *Communications of the ACM, 48*(9).

Kaskalis, T. H., Tzidamis, T. D., Margaritis, K., & Evangelidis, K. (2006). Multimedia creation: An educational approach. *WSEAS Transactions on Information Science and Applications, 3*(2), 470-477.

Koutsakis, P. (2007). Mobile computing symposium: Movement prediction and planning: Integrating latest technology multimedia traffic over high-speed cellular networks. *Proceedings of the 2007 International Conference on Wireless Communications and Mobile Computing IWCMC '07.*

Kulkarni, P., & Öztürk, Y. (2007). Requirements and design spaces of mobile medical care. *ACM SIGMOBILE Mobile Computing and Communications Review, 11*(3).

Kyriakidou, A., Karelos, N., & Delis, A. (2005). Time- and power-sensitive techniques: Video-streaming for fast moving users in 3G mobile networks. *Proceedings of the 4th ACM International Workshop on Data Engineering for Wireless and Mobile Access MobiDE.*

Lanuzza, M., Margala, M., & Corsonello, P. (2005). Special purpose processing: Cost-effective low-power processor-in-memory-based reconfigurable datapath for multimedia applications. *Proceedings of the 2005 International Symposium on Low Power Electronics and Design ISLPED '05.*

Li, J., Chang, S.-F., Lesk, M., Lienhart, R., Luo, J., & Smeulders, A. W. M. (2007). New challenges in multimedia research for the increasingly connected and fast growing digital society. *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval MIR '07.*

Lin, E. T., Eskicioglu, A. M., Lagendijk, R. L., & Delp, E. J. (2005). Advances in digital video content protection. *IEEE: Special Issue on Advances in Video Coding and Delivery, 93*(1), 171-183.

Magerkurth, C., Engelke, T., & Memisoglu, M. (2004). Augmenting the virtual domain with physical and social elements: Towards a paradigm shift in computer entertainment technology. *Computers in Entertainment (CIE), 2*(4).

Magerkurth, C., Engelke, T., & Grollman, D. (2006). A component based architecture for distributed, pervasive gaming applications. *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology ACE '06.*

Miguel, T. C., Barracel, O. F., & Panyella, O. G. (2004). Tools development for arts research and practice: iGlue.v3: An electronics metaphor for multimedia technologies integration. *Proceedings of the 12th annual ACM International Conference on Multimedia MULTIMEDIA.*

Mohammed, N., Abdelhakim, A., & Shirmohammadi, S. (2007). A Web-based group decision support system for the selection and evaluation of educational multimedia. *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education Emme '07.*

Naef, M., Staadt, O., & Gross, M. (2004). Blue-c API: A multimedia and 3D video enhanced toolkit for collaborative VR and telepresence. *Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry VRCAI '04.*

Nang, J., & Park, J. (2007). Database theory, technology, and applications. An efficient indexing structure for content based multimedia retrieval with relevance feedback. *Proceedings of the 2007 ACM Symposium on Applied Computing SAC '07.*

Pereira, F. (2007). MPEG multimedia standards: Evolution and future developments (Tutorial). *Proceedings of the 15th International Conference on Multimedia MULTIMEDIA.*

Petridis, K. B. S., Saathoff, C., Simou, N., Dasiopoulou, S., et al. (2006). Knowledge representation and semantic annotation of multimedia content. *IEE Proceedings—Vision, Image and Signal Processing, 153*(3), 255-262.

Pham, B., & Wong, O. (2004). Computer human interface: Handheld devices for applications using dynamic multimedia data. *Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia GRAPHITE.*

Roccetti, M., & Syed, R. M. (2003). *Proceedings of the International Conference on Simulation and Multimedia in Engineering Education*. Orlando, Florida, USA: The Society for Modeling and Simulation International.

Rowe, L., & Jain, R. (2005). ACM SIGMM Retreat Report. *ACM Trans. Multimedia Computing, Communications, and Applications, 1*(1), 3-13.

Rowe, L. A., & Jain, R. (2005). ACM SIGMM retreat report on future directions in multimedia research. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 1*(1).

Schonberg, D., & Kirovski, D. (2004). Watermarking and multi-media processing: Fingerprinting and forensic analysis of multimedia. *Proceedings of the 12th Annual ACM International Conference on Multimedia MULTIMEDIA'04.*

Sebe, N., & Tian, Q. (2007). Personalized multimedia information retrieval: Personalized multimedia retrieval: the new trend? *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval MIR '07.*

Sher, M., & Magedanz, T. (2006). Security issues for wireless networks: Secure access to IP multimedia services using generic bootstrapping architecture (GBA) for 3G & beyond mobile networks. *Proceedings of the 2nd ACM International Workshop on Quality of Service & Security for Wireless and Mobile Networks Q2SWinet '06.*

Socek, D., Sramka, M., Marques, O., & C´ulibrk, D. (2006). An improvement to a biometric based multimedia content protection scheme. *Proceeding of the 8th Workshop on Multimedia and Security MM&Sec'06,* Geneva, Switzerland.

Sorkin, S., Tupper, D., & Harmeyer, K. (2005). Instructional multimedia institutes for mathematics, science and technology educators. *Journal of Computing Sciences in Colleges, 20*(3).

Syed, M. R., & Baiocchi, O. R. (2001). Intelligent multimedia, computing and communications—Technologies of the future. *Proceedings of ICIMADE'01: International Conference on Intelligent Multimedia and Distance Education*, Fargo, USA (285 pages). John Wiley & Sons, Inc.

Syed, M. R. (2001). *Design and management of multimedia information systems: Opportunities and challenges.* Hershey, PA: Idea Group Publishing.

Syed, M. R. (2002). *Multimedia networking: Technology, management and applications*. Hershey, PA: Idea Group Publishing.

Syed, R. M., & Tareski, V. (2001). Advances in educational technologies—Multimedia, WWW and distance education. *Proceedings of ICIMADE'01: International Conference on Intelligent Multimedia and Distance Education*, Fargo, USA (225 pages). John Wiley & Sons, Inc.

Thwaites, H. (2006). Human computer interaction: Cyberanthropology of mobility. *Proceedings of the 3rd International Conference on Mobile Technology, Applications & Systems Mobility '06.*

Uludag, U., Pankanti, S., Prabhakar, S., & Jain, A. (2004). Biometric cryptosystems: Issues and challenges. *IEEE Special Issue on Enabling Security Technologies for Digital Rights Management*, *92*(6), 948-960.

Verkasalo, H. (2006). Empirical observations on the emergence of mobile multimedia services and applications in the U.S. and Europe. *Proceedings of the 5th International Conference on Mobile and Ubiquitous Multimedia MUM '06.*

Viljoen, D. W., Calitz, A. P., & Cowley, N. L. O. (2003). A 2-D MPEG-4 multimedia authoring tool. *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa AFRIGRAPH.*

Wei, B., Renger, B., Chen, Y. F., Jana, R., Huang, H., Begeja, L., Gibbon, D., Liu, Z., & Shahraray, B. (2005). Applications on the go: Media Alert—A broadcast video monitoring and alerting system for mobile users. *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services MobiSys*.

Yang, J., Li, Q., & Zhuang, Y. (2002). Multimedia: OCTOPUS: Aggressive search of multi-modality data using multifaceted knowledge base. *Proceedings of the 11th International Conference on World Wide Web WWW '02.*

Yi, X., Tan, C., Siew, C., & Syed, M. R. (2002). ID-based agreement for multimedia encryption. *IEEE Transaction on Consumer Electronics, 48*(2).

Yu, C. C. (2004). Innovation, management & strategy: A web-based consumer-oriented intelligent decision support system for personalized e-services. *Proceedings of the 6th International Conference on Electronic Commerce ICEC.*

Yu, H. H., Kundur, D., & Lin, C. Y. (2001). Spies, thieves, and lies: the battle for multimedia in the digital era. *IEEE Multimedia, 8*(3), 8-12.

Zhiwen, Y., Xingshe, Z., & Daqing, Z. (2005). An adaptive in-vehicle multimedia recommender for group users. *Vehicular Technology Conference, VTC 2005—Spring 2005 IEEE 61* (pp. 2800-2804, Volume 5, Issue 30).

Section 1

Fundamental Concepts and Theories

This section serves as the foundation for this exhaustive reference tool by addressing crucial theories essential to the understanding of multimedia technologies. Chapters found within these pages provide an excellent framework in which to position multimedia technologies within the field of information science and technology. Individual contributions provide overviews of multimedia education, multimedia messaging, and multimedia databases while also exploring critical stumbling blocks of this field. Within this introductory section, the reader can learn and choose from a compendium of expert research on the elemental theories underscoring the research and application of multimedia technologies.

Chapter 1.1
Fundamentals of Multimedia

Palmer W. Agnew
State University of New York at Binghamton, USA

Anne S. Kellerman
State University of New York at Binghamton, USA

ABSTRACT

This chapter introduces multimedia, defined as interacting with information that employs most or all of the media: text, graphics, images, audio, and video. Students and faculty need to learn to create and use high-quality multimedia documents, including references, lecture materials, reports, and term papers. The authors provide a framework for understanding multimedia in its rapidly changing context. They discuss a wide spectrum of multimedia end-user devices that range from smart cell phones and powerful PCs to intelligent cars and homes. They also propose a vision of pervasive multimedia any time and anyplace, and discuss related issues, controversies, and problems. Typical problems are excessive complexity and a plethora of choices that paralyze many potential users. The chapter concludes with a discussion of possible solutions to major problems and probable future trends.

INTRODUCTION

Multimedia is interacting with text, graphics, images, audio, and video. Creators and users of multimedia employ end-user devices that range from PCs and interactive televisions to smart phones and PDAs. People exchange multimedia using delivery methods such as dial-up and cable-modem access to the Internet, mailed DVDs, and Internet2. Multimedia communications can be more effective and interesting than communications that are limited to text.

Most of us will create, as well as use, multimedia throughout the remainder of our lives. Almost all future work and everyday life will involve dealing with multimedia wherever we are by using the end-user devices at hand. Examples of use include sending images to Aunt Lizzie by way of a cell phone, and writing and wirelessly posting a report on the Internet concerning worldwide petroleum sources, while standing near an oil well in the Middle East.

The objectives of this chapter are to:

- Provide a framework for efficiently acquiring new knowledge and skills in the rapidly changing multimedia arena;
- Discuss a vision for effective multimedia creation and use, by nearly everybody, nearly anywhere, and at any time;
- Delineate the major issues, controversies, and problems that litter the path toward achieving that vision; and
- Discuss some corresponding solutions and recommendations, many of which involve skills that instructional technologists, teachers, and students need.

BACKGROUND

Figure 1 shows a high-level view of people and components involved in creating and using multimedia. Some providers create multimedia content, information, titles, or applications that employ multiple media and are interactive. These authors create this content by employing end-user device hardware and software. They then typically store the resulting content on servers. The content is delivered to other, often different end-user devices employed by users, customers, or readers by means of delivery networks that range from mailing diskettes or DVDs to using local area networks or the Internet. By no means are all authors professional creators; the most interesting aspect of multimedia is that it is now sufficiently inexpensive and almost sufficiently easy that almost anybody can create as well as use multimedia. Other providers include a wide range of individuals, companies, and governments that play a wide variety of roles. For example, some providers provide products and services that are important to the delivery of multimedia to end-users.

Figure 1 is a framework in which you can add newly acquired knowledge about multimedia. For example, if you think you might want to provide an on-demand multimedia tutorial for your students to use from their cell phones, you need to have an authoring end-user device, a way to deliver this content to your users, and tools to allow you to create content for the desired end-user platform. You should know that, at least in the US, wireless delivery will be problematic. The good news is that network providers are improving their products and services, with the goal of handling wireless high-quality images and video within the next couple of years.

Creators and end-users employ a variety of tools. Some tools operate on the individual media.

Figure 1. Framework

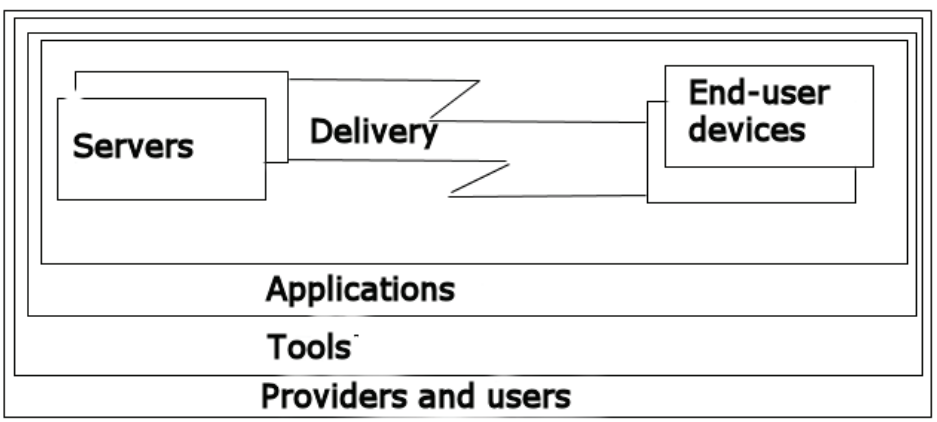


Table 1. Typical multimedia software

Medium	Tools
Text	Microsoft Word, Corel WordPerfect, Tex, Latex
Graphics (i.e., vectors)	Corel Draw, Adobe Illustrator, Macromedia Fireworks, Adobe ImageReady, Macromedia Flash
Image (i.e., bitmap)	Adobe Photoshop, Jasc Paint Shop Pro, Macromedia Fireworks
Audio	Sony Sound Edit Pro, Sony Sound Forge for Windows, Sony Acid, Cakewalk products
Synthetic Video (animation)	AutoDesk AutoCAD, Discreet 3D Studio (MAX), Virtus 3D Website Builder, Macromedia Flash, Electric Image Amorphium Pro, Alias Maya
Captured Video	Adobe Premiere, Avid, Media 100 products, Ulead Media Studio Pro, Microsoft MovieMaker, Apple iMovie
Authoring Systems for All Media	Macromedia Director, Macromedia Dreamweaver, Click2learn Toolbook, Microsoft Front Page, Adobe Page Mill, Microsoft PowerPoint with Producer

Other tools, called multimedia authoring systems, assemble multimedia media and add interactivity. Individual media tools include editors for text, images, graphics, audio, and video. Tools that generate HTML with associated Web browsers are examples of multimedia authoring systems. Table 1 provides some specific examples. A great many can be found at Maricopa Community College's Web site, Multimedia Authoring (Maricopa Community College, 2004).

One correct conclusion that you can draw from the table's examples is that creators of multimedia often need to obtain and master separate tools for each medium, as well as one or more authoring tools that combine several media. To the extent that multimedia creators and users are different people, users are also faced with multiple choices of tools with which to play back the media, such as browser plug-ins for a variety of streaming audio and video formats. Users may also need one or more authoring systems' so-called "run-time environments," if not the full authoring systems.

Multimedia tends to be large in terms of storage space and network bandwidth. Whereas a text page occupies only about 3,000 bytes, five minutes of high-quality video can occupy up to 1 Gigabyte (1 GB). Solutions involve squeezing multimedia by compressing it using a range of sophisticated technologies. As you can see from Table 2, desktop computers routinely come with 80 GB of hard-drive storage space, whereas mobile devices have far less space. Multimedia requires networks that are fast enough to ensure that multimedia plays correctly. Cable modems, Digital Subscriber Lines, and special techniques of streaming are required and often perform adequately, but note the typical network data rate of PDAs in Table 2.

Assume that you have set up an environment such as the one in Figure 1. Now comes the really

Table 2. Typical end-user devices

Property	Desktop Computer	PDA and Cell Phones
Storage	80 GB hard drive	128 MB Flash RAM card
Processor clock rate	2.4 GHz	400 MHz
Operating system	Windows XP	PocketPC
Display diagonal, inches	17.0	3.8
Network for delivery	Cable modem, wireline, typical speed of 400 K bits/sec	Wireless, typical speed of 16 Kbits/sec

Table 3. Sample content-preparation guidelines

Guideline
Provide user-selectable options for multiple learning styles.
If you present text and audio on the same page, make sure that both use the same words. Nobody can read one set of words while listening to different words.
Remember that any movement on a page, such as video or animation, strongly distracts a user from non-moving information such as text.
Use colors to facilitate readability, appeal to different audiences, and reflect different cultures.
Use background music sparingly, to set a mood.
Prepare text for skimming rather than for reading long passages in detail. Computer monitors tire the eyes because they have resolutions that are much less than resolutions of printed pages.
Organize multimedia information carefully and consistently.
Have important information no more than three clicks (or levels) deep.
Do not use copyrighted media in any content that you sell, unless you obtain the owner's permission. Be careful what copyrighted media you use, even in an educational context.

For extensive details, consult *Distributed Multimedia* (Agnew & Kellerman, 2002).

challenging part: creating the content or coaching others to create content. If you are coaching students, you should know that creating multimedia projects can help your students to achieve goals that include developing higher-order thinking skills and interpersonal skills, learning content by engaging multidisciplinary subjects, and developing technical competence and media literacy that will empower the students throughout their lives (TeleEducation NB, 2002). Furthermore, employing multimedia will allow you to improve students' efforts outside formal classroom settings and allow you to piggyback on these efforts inside the classroom. Victory, however, is not merely receiving any old PowerPoint presentation, even with audio, animation, and video. Having an

arsenal of past examples and setting up Olympic scoring of students' results (Agnew, Kellerman, & Meyer, 1996) helps elevate the standards of content quality. Guidelines such as those in Table 3 can help you as either a coach or creator.

ISSUES, CONTROVERSIES, AND PROBLEMS

Early adopters and pioneers routinely create multimedia. Today, with affordable hardware and software, people who are not media professionals can prepare interactive documents that contain multiple media. They can communicate using these media either in real time, using streaming, or asynchronously, using e-mail. Using multimedia tools, they can structure their information to greatly facilitate understanding. They can send this multimedia information to remote locations, with some certainty that it will be received rapidly and understood successfully. Tomorrow, lacking such skills will seriously impact almost all professionals' and students' lives, job satisfaction, and job performance. Content created by early adopters and pioneers has made multimedia literacy part of the expected price of admission to many roles.

Complexity and too many choices stymie the rest of us. Unfortunately, arcane skills are still required to set up digital environments and then use them for serious information preparation or even for living-room enjoyment.

Our professional colleagues and teaching friends confirm that multimedia is changing fast and furiously, making it difficult for them to keep up in a field that they view as diverging from their respective professional areas. They are elementary school teachers, biology teachers, instructional technologists, literature professors, and so forth. Nevertheless, new multimedia capabilities are superb ones that such professionals need. What best characterizes humans, all over the world, is our ability to communicate orally and in writing. Multimedia can make this communication even better, if done well.

Dramatic price declines in software and hardware allow many more of us to set up the environments we need to both create and use multimedia. For example, we have watched PC cams (small video cameras that connect to PCs for both storage and control) drop in price below $100, then below $30, and now below $10. The difficulty all of us find is that to learn what we specifically need to know, we must either spend hours on the Internet or buy books, often blindly. If only the cost and time spent learning to use the affordable hardware and software were decreasing as well.

Vendors have conflicting motivations. It would be easier for users if Microsoft could indeed integrate all required support into its operating system. Governments and competing vendors have widely divergent motivations. As a result, authors are required to either guess at which formats and data rates their audience requires or supply several different views of the same content.

Book stores display many shelves of books on different media and on enhancing Web pages. You are led to believe, with some truth, that to learn to create multimedia, you need a yard of books. Very few books tell how to create meaningful content where content is targeted to academic or serious professional uses.

Mismatch of equipment between home and school. People are using digital cameras, both standalone and inside cell phones, in record numbers. They are learning how to use them on their own, just blundering through. Electronic gadgets including computers at home are proliferating. Home computers tend to be good ones that far outperform what is available in most K-12 schools and many universities. In school, children are exposed to old equipment without getting sufficient quality instruction in its use and without educators taking advantage of the motivational aspects of using the equipment to enhance education. The Salvation Army prefers

to reject any computer that is more than three years old; can schools do the same?

Creating high-quality content is difficult today, and few are experienced in it. The available tool often drives the content, rather than conversely. A tragically famous example of this is NASA's chart justifying its tests of foam striking a Space Shuttle's wing. The authors produced bullet after bullet of PowerPoint text at random levels of hierarchy and importance. Unfortunately, they buried the only important fact (that their test set-up had been totally unrelated to the real world) in a sentence fragment at the bottom of the page, rather than making it the title, or at least phrasing it as an unambiguous complete sentence.

Dr. Edward Ayers (2004) understands that few scholars have done what he has done over the last decade, namely scholarly research on multimedia narration (Ayers, 2004; valley.vedh.virginia.edu).

SOLUTIONS AND RECOMMENDATIONS

Early adopters and pioneers routinely create multimedia. To help students and teachers achieve multimedia literacy as part of life's expected price of admission, educational technologists must continue learning, coaching, and mentoring the rest of the academic community. Academics in universities must work together. As an example of the kinds of successes cooperation can have, Dr. Meyer (a computer science professor) and Dr. Driver (a literature professor) worked together to inspire literature students to compose multimedia documents. Each provided relevant expertise (Driver & Meyer, 1999).

Young people are intrinsically early adopters and pioneers. While it is charming when students teach teachers, it is teachers who can best apply the required pedagogy to achieve the desired results.

Complexity and too many choices stymie the rest of us. A viable approach is to identify a small set of tools and stick with it for several years, ignoring incremental advances in hardware and software. This includes taking advantage of tools that ship with major software. For example, both Microsoft and Apple ship moderately functional video-editing software as parts of their current flagship operating systems.

Mismatch of equipment between home and school. Parents need to get involved, working with teachers to encourage home equipment use for creation of rich media content in support of pedagogical goals.

Creating high-quality content is difficult today, and few are experienced in it. As it becomes easier to assign multimedia, this will become more common. Future teachers need coaching in how to use multimedia in the classroom, as in the text, *Multimedia in the Classroom* (Agnew et al., 1996). Teaching teachers in colleges and universities does not consist of imparting isolated media skills, but rather of showing how to put it all together in a learning environment. One key is exhibiting exemplary multimedia content achieved by relevant people. It is not useful to present George Lucas's work for students to emulate; work produced by students in prior years is far more helpful. Like many subjects, the hardest part of teaching multimedia is getting started.

Teachers need to develop pedagogical goals for both teaching basic multimedia literacy and using multimedia to teach other subjects. This requires that they understand what is required to create and use multimedia to achieve pedagogical goals and that they then implement paths to reaching these goals that are consistent with their existing environments. They must be able to assess results along multiple dimensions. Merely using multimedia is not enough when the goal is quality content supported by rigorous arguments. Teachers and educational technologists will routinely need to advise, recommend, approve, and obtain parts of the framework required for multimedia.

Educators will need to improve their multimedia creation and playback environments, as prices and quality of tools continue to decline, and in step with the mainstream and with overall mission objectives. For example, the military now uses high-speed computers, high-bandwidth delivery connections, and sophisticated multimedia simulations to instruct soldiers on how to carry out missions. Whereas K-12 teachers might have more modest requirements, everyone involved in multimedia must learn to communicate intelligently and efficiently with vendors of multimedia products and services. A major part of this is learning when to search the Web and when to telephone a help line. Experience will assist educators in knowing what multimedia they and students can produce and use with what they have on hand, and what multimedia is suitable for what their intended users will have on hand.

Everyone should be able to recognize and produce quality multimedia content. Those who can do this will need to be generous with their time and talents, and help others to do likewise. Available tools must not drive content in undesirable directions, such as PowerPoint's leading from sentences to sentence fragments to sloppy thinking. Using tools must not be allowed to become the end in itself.

Suitable goals include preparing teachers and students who are routinely able to create and deliver a multimedia proposal, resume, presentation, set of instructions, or report using appropriate selections from alternative media. In many cases the media will need to be suitable for nationwide or even worldwide distribution over the Internet. This requires some understanding of the available bandwidths of connections that range from reasonably fast broadband to abysmally slow cellular wireless. It also requires selection between real-time and asynchronous communication and suitability for a wide variety of end-user devices.

Creating multimedia requires people to be able to:

- Acquire and prepare images and graphics for inclusion in documents;
- Repurpose slides and analog photos into digital forms for playback on TVs, PCs, cell phones, and PDAs;
- Create and incorporate video using either economical PC cams or high-quality Mini-DV camcorders;
- Prepare animation for illustration and instruction; and
- Employ voice recognition software for communicating with devices that are too small to have workable keyboards.

Examples of specific skills include:

- Converting media from one form to another for use in various situations and on various devices;
- Using tools that allow processing of several media files in one batch, such as in converting an entire folder of several images from one format to another at once; and
- Organizing raw and completed information for effective later re-use and collaborative use.

All of these skills require particular attention to what is actually feasible to achieve meaningful results.

Moreover, everyone must know about the rights that educators and others have with respect to using media created by others. Rights may differ between content used as is and altered content.

FUTURE TRENDS

For better or for worse, future uses of multimedia in educational and other contexts will be driven by improvements in technologies. Moore's Law famously predicts that processor power and internal memory capacity will continue to double every year or two, without significant increases in

cost. Less famously, other important technologies are improving at even faster rates. A decade ago, the available hard drive space on a typical home computer sufficed for storing barely 30 minutes of moderate-quality video; now a typical computer can store about 30 hours of high-quality video. Increasingly, economical hard drives have allowed many users to afford personal video recorders, such as TiVo, as well as sophisticated in-home video editing. A decade ago, transcontinental telephone connections were so scarce and expensive that long-distance telephony was a cash cow; now a glut of worldwide fiber-optic cable bandwidth allows Dell to route help-line calls to Indonesia as cheaply as to Indiana.

Companies including Intel, IBM, and AT&T are severely threatened by rapid technological progress that gives customers the opportunity to pay vastly fewer dollars for the same products and services. Many companies see multimedia as their only hope for convincing customers to purchase far more products and services, for at least the same number of dollars. It works. Most owners of digital still cameras now use more hard-drive space for storing images than they used for all purposes just a few years ago. New, tiny video cameras that temporarily store video on expensive flash RAM cards will permanently fill up even more storage space on hard drives.

This trend toward rapid improvement in technologies means four things to us all, as creators and users of multimedia. First, we must beware of pressures to use more of the products and services that do not benefit us, merely because they are increasingly economical. Second, we must try to make excellent use of newly affordable methods for achieving our goals. Third, we must encourage providers of multimedia products and services to make major improvements in ease of learning and ease of use, rather than merely dumping in more and more capabilities with attendant increases in confusion. Fourth, we must continually keep abreast of research on assessments of multimedia technologies, especially ones which include best practices. One way to keep abreast is to subscribe to Kotlas's newsletter, *CIT Infobits* (2004) and review the list she compiled on Assessments of Multimedia Technology in Education.

CONCLUSION

We will all communicate using technologies that range from today's e-mail and Word documents to tomorrow's video instant chats. Each communication will be significantly enhanced by our ability to easily and effortlessly use appropriate media and to allow the recipient to selectively access content in a variety of forms. We will build on skills acquired at home, in K-12 schools, and in higher education. We can expect to learn continually as technologies improve. As vendors expect to sell more of their hardware and software, we hope their motivations will include making equipment easier and quicker to learn and use, as well as more economical to buy. We must help one another, even if it requires cross-disciplinary activities. The best authors need to take advantage of multimedia to create multimedia and to help the rest of us learn to create it and use it.

REFERENCES

Agnew, P., Kellerman, A., & Meyer, J. (1996). *Multimedia in the classroom*. Needham Heights, MA: Allyn & Bacon.

Agnew, P. & Kellerman A. (2002). *Distributed multimedia—technologies, applications, and opportunities in the digital information industry, a guide for users and providers* (2nd edition). Cincinnati: Atomic Dog Publishing.

Assessments of Multimedia Technology in Education: Bibliography. (Information Resource Guides Series #IRG-11), Institute for Academic Technology (IAT). Compiled by Carolyn Kotlas, MSLS, University of North Carolina. This docu-

ment includes government, private industry, and K-12. Retrieved February 10, 2004, from *www.unc.edu/cit/guides/irg-11.html*

Ayers, E. (2004). Doing scholarship on the Web: 10 years of triumphs—and a disappointment. *The Chronicle of Higher Education,* (January 30), B24-B25.

Driver, M. & Meyer, J. (1999). *Engaging students in literature and composition using Web research & student constructed Web projects.* Retrieved February 10, 2004, from *csis.pace.edu/~meyer/hawaii/*

Educase. (2003). *Information resources library on articles with multimedia applications and campus learning.* Retrieved February 10, 2004, from *www.educause.edu/asp/doclib/subject_docs.asp?Term_ID=364*

Koltas, C. (2004). *CIT infobits.* Retrieved February 10, 2004, from *www.unc.edu/cit/infobits/index.html*

Maricopa Community College. (2004). *Multimedia authoring.* Retrieved February 10, 2004, from *www.mcli.dist.maricopa.edu/authoring/*

TeleEducation NB. (2002). *The significant difference phenomenon.* Retrieved February 10, 2004, from *teleeducation.nb.ca/significantdifference/*

This work was previously published in Technology Literacy Applications in Learning Environments, edited by D. D. Carbonara, pp. 263-273, copyright 2005 by Information Science Publishing (an imprint of IGI Global).

Chapter 1.2
Digital Multimedia

Neil C. Rowe
U.S. Naval Postgraduate School, USA

INTRODUCTION

Multimedia data can be important assets of government computer systems. Multimedia data can be documents, statistics, photographs and graphics, presentations, video and audio of events, and software. Examples include maps, video of meetings, slide presentations by consultants and vendors, graphs of budgets, and text of regulations. Video of meetings of legislatures and other government organizations is particularly valuable as it makes government processes more visible to citizens and can encourage trust in government. Multimedia is also particularly valuable in presenting geographical information (Greene, 2001), a concern of all governments. Added multimedia can also be used to more effectively deliver information to people, as with films, animations, sound effects, and motivational materials.

Multimedia information is important for digital government because it is often a more natural communication mode for people than text. It is thus important that government be responsive to the needs and desires of citizens by providing it. Much of the world is illiterate, and the ubiquity of television means even the literate often prefer watching video to reading text. Some citizens have special needs: Blind people need audio, and deaf people need images. Video and audio also convey information beyond text: A video of a legislature meeting contains subtleties not apparent from its transcript. Research has shown that multimedia is especially good at conveying explanatory information about functional relationships in organizations (Lim & Benbasat, 2002). Research has also shown that people learn better from multimedia presentations than from conventional classroom instruction, and the multimedia provides a consistent experience available at any time unlike human instructors (Wright, 1993).

BACKGROUND

The management of multimedia data entails considerations not encountered with data that is solely text (Vaughan, 2003). The main problem is data volume: A typical report can be stored in 20,000 bytes, but a typical 20-centimeter square image requires 500,000 bytes to represent adequately, an audio clip that is 1 minute long requires around 1,000,000 bytes, and a typical 3-centimeter square

video clip that is 1 minute long requires around 50,000,000 bytes. Compression techniques can reduce storage requirements somewhat; however, media that can be compressed significantly tend to be merely decorative and not very useful for digital government. Multimedia size is especially a problem when transferring media between computers, especially with the limited data rates of conventional telephone lines and modems (Rao, Bojkovic, & Milovanovic, 2002). So, since digital government cannot be sure what technology its citizens have, it must be conservative in its use of multimedia.

Distributed database technology (Grosky, 1997) can help manage multimedia data efficiently. However, the human side of multimedia management requires a different set of skills than those of most computer support staff. One needs media specialists familiar with the problems of the technology, including some staff with art training to address the aesthetic issues that arise. Much multimedia management is time consuming, so adequate personnel must be available. Government can also choose to actively encourage the development of a multimedia-supporting infrastructure by its industries (Mohan, Omar, & Aziz, 2002).

INDEXING AND RETRIEVAL OF MULTIMEDIA

A first issue in using multimedia data is finding it. Citizens often want to retrieve quite specific multimedia objects to meet their needs, and they can do this with a browser or a search engine (Kherfi, Ziou, & Bernardi, 2004). This requires metadata describing the media objects such as size, date, source, format, and descriptive keywords. A browser can provide a hierarchy of media objects that users can navigate. This works best when media objects can be described in just a few words, or are characterized along a single dimension like date or place. Otherwise, a keyword-based search engine is necessary, such as that provided by commercial services like Google but adapted to search only government data. Accommodating a broad range of citizens means keeping extensive synonym lists for keyword lookup so that many possible ways of specifying the same thing will work. In some cases, a graphical specification may be a good way for the public to specify what they are looking for, such as a visual timeline or geographical display on which they click on the location they want.

Unfortunately, it is difficult to index and search nontext media for its contents. Segmentation by identifying shapes within images can be tried, but it is time consuming and often unreliable. So captions on media objects (text that describes them) are valuable (Rowe, 2005). They can be directly attached to a media object or located near it in a document. Captions directly attached include the name of the media file, descriptive keywords, annotations describing the object like the *alt* string associated with Web media, the text of clickable links on the Web that retrieve the object, text directly embedded in the media like characters drawn on an image, and data in different channels of the media like narration or closed captions for video. Captions indirectly attached include text centered or in a special font above or below the media, titles and section headings of the document containing the media, special linguistic structures like "The photo shows …," and paragraphs above or below the media. Caption text can be indexed for a more precise keyword search than that obtained by just indexing all the words of the enclosing document (Arms, 1999). This is what the media search engines such as the image searchers of Google and AltaVista do, though their success rate at finding images is not as good as their success rate at text search. Media retrieval is, however, an active area of research, and new developments are appearing frequently.

DELIVERY OF MULTIMEDIA

Multimedia can enhance documents in many ways. It can enliven information, and this is helpful for the often-unexciting information provided by governments. But the primary concern of government must be for media that convey important information of their own. Mostly, this means the delivery of multimedia information from the government to its citizens, though there are also issues in the collection of information by government (Cheng, Chou, Golubchik, Khuller, & Samet, 2001) and communications within government.

Broadcast technology is the traditional method for a government to disseminate media through newspapers, radio, and television. Broadcast is a one-way technology. This is fine for announcements and authoritarian governments, but interactivity is very important to a responsive and effective government. So, the Internet and especially the World Wide Web are increasingly preferred to deliver user-selected information and permit the completion of forms. The Web is well suited for multimedia. It permits the embedding of pointers to multimedia content in documents with much the same ease as text. Web multimedia can range from informal illustrations to media retrieved from structured databases in response to queries entered on dynamic Web pages. Media is particularly helpful for the illiterate as graphical interfaces can enable access to the full power of computers without the necessity of words.

A key issue for digital government is the choice of media formats. Government information systems intended for only internal use can follow very few mandated formats for interoperability. But much multimedia is for the public, and the public uses a diversity of computers, software, and networking services, and accessibility to all citizens is important. So copies of important multimedia in different formats are essential. Web images are currently mostly in JPEG and GIF formats, with some in PNG format. Audio and video are more diverse: Currently popular audio formats are WAV, RAM, MID, and ASX, and currently popular video formats are MPEG, SWF, MOV, FLI, AVI, and MP3. Multimedia can also be software in various formats. Not all these formats are supported by all Web browsers, so it is important for government to provide free viewer software for downloading so citizens can view any of its multimedia. This generally restricts governments to using formats that have free open-source software for reading or viewing them.

Multimedia software of particular interest to government organizations is groupware, supporting collaborative activities of groups. It can be used to run meetings of people at widely scattered locations so participants can see, hear, and read comments by other participants. Groupware requires transmission of video, audio, and graphics between sites (Klockner, Mambrey, Printz, & Schlenkamp, 1997).

STREAMING MULTIMEDIA

Because of its bulkiness, multimedia is often best stored at a few centralized sites and retrieved from there. That can entail logistical problems since video and audio in particular need to be delivered at a certain minimum speed to be viewable or listenable (Smith, Mohan, & Li, 1999). Video or audio can be fully downloaded in advance of playing, but this entails a time delay and most citizens prefer real-time delivery (streaming). Important applications with streaming are being accomplished including video meetings, video medicine, distance learning, multimedia mail, and interactive television. Streaming is simplified if it is delivered by one-way broadcast, and this works well for standardized content such as training materials. Traditional technology like television can also be effective in the streaming of government media (Morisse, Cortes, & Luling, 1999) but requires citizens to access it at a particular time. Another alternative is to supply

citizens with an optical disk (CD-ROM or DVD) containing the media.

The biggest challenge with streaming is ensuring adequate bandwidth (rate of data transmission). A single MPEG-1 compressed video of standard television-picture quality needs around 2 megabits per second (though videoconferencing and speeches can be adequate with less), and music audio requires 1 megabit per second. Standard (ISDN) telephone lines provide 0.064 megabits per second, inadequate for both. T-1, T-2, and other lines can improve this to theoretically 1.5 megabits per second, but that is still inadequate for video. Digital subscriber lines using cable-television technology can provide higher bandwidths using network technologies such as ATM, but even those can be pressed to produce real-time video. Data-compression techniques can reduce bandwidth somewhat; typical maximum compression ratios range from 2:1 for audio to 20:1 for images and 50:1 for video. Other tricks with noncritical video are to periodically skip frames or decrease the size of the images.

Transmission bandwidth is also bounded by the delivery rate of the media from an archive. Live video may be able to bypass storage and go directly onto the network. But in general, multimedia data must be archived in blocks for efficient memory management, though the blocks can be larger than those typical for text. Magnetic tape and optical disks are less flexible in manipulating blocks than magnetic disks, so the latter is preferable for multimedia. But it does take time for a disk head to go between blocks, so successive blocks can be put on different disks so that a block from one disk can be transmitted while the other block is being readied (striping).

Besides bandwidth limits, networks can also have transmission delays (latency), which affects real-time interactive applications like videoconferencing. Delays can be minimized by good network routing (Ali & Ghafoor, 2000; Gao, Zhang, & Towsley, 2003). Using different paths to relay different parts of multimedia data from source to destination can reduce the effect of any one bottleneck. If data loss is especially important to avoid, redundant data can be sent over the multiple paths. But much video playback can tolerate occasional data loss.

Another transmission issue is the evenness of the arrival of multimedia data at a destination site (burstiness or jitter) since unevenness can cause unacceptable starts and stops in audio or video. This can happen when other network traffic suddenly increases or decreases significantly. Usually, video delays cannot exceed 0.1 seconds and music-audio delays 0.0001 seconds, so this is a key quality-of-service issue. Caching of data in storage buffers at the delivery site reduces the problem, but effective multimedia buffers must be big. Transmission by multiple routes will also help.

Multimedia delivery is still more difficult when the destination device is a small handheld one since these are limited in memory, processor speed, and networking capabilities. Although multimedia is better displayed on conventional computer hardware, many users prefer such small devices to access the Internet while engaged in other activities. Then streaming is necessary and must be done with significant bandwidth and screen-size limits. Managing the display of information on handheld devices is called content repurposing and is an active area of research (Alwan et al, 1996; Singh, 2004).

Other important technical issues in streaming include the following (Jadav & Choudhary, 1995).

- System architectures should be chosen for fast input and output; parallel ports are desirable.
- Star and fully connected network topologies are desirable. That may only be feasible with local networks for many applications.
- Switches should be preferred to routers on network connections since the routers have lower bandwidths and higher delays.

- Experimentation with the packet size for multimedia data may improve performance since the best size is hard to predict.
- Caching of frequently used data can help efficiency since some media objects will be much more popular than others.

Thus, without careful design there can be serious problems in multimedia delivery. These problems can be negotiated between senders and receivers, either beforehand or at the time of transmission. Generally speaking, worst-case or deterministic metrics for quality of service are more important than average-case or statistical ones for real-time multimedia since users have limits in unacceptable speed, delay, or jitter.

ARCHIVING OF MULTIMEDIA

Governments must archive many important records, and this includes multimedia information. The size of media objects necessitates high storage costs and slow access. While the costs of storage media continue to decrease, now more data is being created in the first place.

A problem for archiving is the diversity of storage devices and media formats. New ones arise continually, so archiving requires either archiving the hardware and software that can read the stored media even when the hardware, software, and formats are no longer being used for new media, or else periodically recopying the data into new devices and formats (Friedlander, 2002). For instance, the Library of Congress of the United States still has audio stored on wire media, a technology long obsolete. Video may be a particular problem in the future as there are several competing formats today as well as an ongoing shift to high-definition images. Unfortunately, digital media is different from old books in that it can be virtually undecipherable without the right hardware and software to read it. So governments must continually invest to support their media archives.

Another problem for digital government is to decide what to archive (Liu & Hseng, 2004). Legal requirements may specify particular archiving, but governments have a responsibility to anticipate future data needs as well. Copyright issues can actually simplify as media ages and enters the public domain. But when money for archiving is limited, conflicting interests within the public may disagree as to what information to keep, and politics may be needed to resolve this.

FUTURE TRENDS

The increasing speed of computers and networks and the decreasing cost of digital storage will enable increased use of multimedia in computer systems in the future. This will enable governments to store and offer increasing amounts of multimedia data for their citizens. Media like video of public meetings, census maps, and graphics documenting government practices will become routinely available so that all citizens can access them without needing to be physically present at government facilities. The main challenges are indexing all these media so citizens can find them, and delivering them (particularly video) across a network fast enough to be useful. Speed and cost improvements alone will not solve all the technical problems as other issues discussed above must be addressed, too. It will take a number of years to reach levels of adequate government media service even in technologically advanced countries.

CONCLUSION

Multimedia is a natural way for people to communicate with computers, more natural than text, and should become an increasingly common mode for people to learn about and participate in

their governments. Good planning is necessary, however, because the data sizes of multimedia objects can be significantly larger than those of text data. This means that media delivery can be unacceptably slow or uneven with current technology, and this limits what can be offered to citizens. Nonetheless, the technology to support media access is improving, and governments will be able to exploit this.

REFERENCES

Ali, Z., & Ghafoor, A. (2000). Synchronized delivery of multimedia information over ATM networks. *Communications of the ACM, 43*(11), 239-248.

Alwan, A., Bagrodia, R., Bambos, N., Gerla, M., Kleinrock, L., Short, J., et al. (1996). Adaptive mobile multimedia networks. *IEEE Personal Communications, 3*(2), 34-51.

Arms, L. (1999). Getting the picture: Observations from the Library of Congress on providing access to pictorial images. *Library Trends, 48*(2), 379-409.

Cheng, W., Chou, C., Golubchik, L., Khuller, S., & Samet, H. (2001, May 21-23). Scalable data collection for Internet-based digital government applications. *Proceedings of the First National Conference of Digital Government Research,* Los Angeles, CA (pp. 108-113).

Friedlander, A. (2002). The National Digital Information Infrastructure Preservation Program. *D-Lib Magazine, 8*(4). Retrieved January 13, 2006, from http://www.dlib.org/dlib/april02/friedlander/04friedlander.html/.

Gao, L., Zhang, Z.-L., & Towsley, D. (2003). Proxy-assisted techniques for delivering continuous multimedia streams. *IEEE/ACM Transactions on Networking, 11*(6), 884-894.

Greene, R. (2001). *Opening access: GIS in e-government.* Redlands, CA: ESRI (Environmental Systems Research Institute) Press.

Grosky, W. (1997). Managing multimedia information in database systems. *Communications of the ACM, 43*(12), 72-80.

Jadav, D., & Choudhary, A. (1995). Designing and implementing high-performance media-on-demand servers. *IEEE Parallel and Distributed Technology, 3*(2), 29-39.

Kherfi, M., Ziou, D., & Bernardi, M. (2004). Image retrieval from the World Wide Web: Issues, techniques, and systems. *ACM Computing Surveys, 36*(1), 35-67.

Klockner, K., Mambrey, P., Prinz, W., & Schlenkamp, M. (1997, September 1-4). Multimedia groupware design for a distributed government. *Proceedings of the 2nd EUROMICRO Conference,* Budapest, Hungary (pp. 144-149).

Lim, K., & Benbasat, I. (2002). The influence of multimedia on improving the comprehension of organizational information. *Journal of Management Information Systems, 19*(1), 99-127.

Liu, J.-S., & Hseng, M.-H. (2004). Mediating team work for digital heritage archiving. *Proceedings of International Conference on Digital Libraries* (pp. 259-268).

Mohan, A., Omar, A., & Aziz, K. (2002). Malaysia's multimedia Super Corridor Cluster: Communication linkages among stakeholders in a national system of innovation. *IEEE Transactions on Professional Communications, 45*(4), 265-275.

Morisse, K., Cortes, F., & Luling, R. (1999). Broadband multimedia information service for European parliaments. *Proceedings of International Conference on Multimedia Computing and Systems,* (Vol. 2, pp. 1072-1074).

Rao, K., Bojkovic, Z., & Milovanovic, D. (2002). *Multimedia communication systems: Techniques, standards, and networks.* Englewood Cliffs, NJ: Prentice-Hall.

Rowe, N. (2005). Exploiting captions for Web data mining. In A. Scime (Ed.), *Web mining: Applications and techniques* (pp. 119-144). Hershey, PA: Idea Group Publishing.

Singh, G. (2004). Content repurposing. *IEEE Multimedia, 11*(1), 20-21.

Smith, J., Mohan, R., & Li, C.-S. (1999, October). Scalable multimedia delivery for pervasive computing. *Proceedings of the Seventh ACM Conference on Multimedia,* (Vol. 1, pp. 131-140).

Vaughan, T. (2003). *Multimedia: Making it work* (6th ed.). New York: McGraw-Hill Osborne Media.

Wright, E. (1993). Making the multimedia decision: Strategies for success. *Journal of Instructional Delivery Systems, 7*(1), 15-22.

KEY TERMS

Bandwidth: Amount of data (measured in bits per second) that can be transmitted across a network.

Broadcast: Transmission of some data in one direction only to many recipients.

Caption: Text describing a media object.

Data Compression: Transforming of data so that they require fewer bits to store.

Groupware: Software to support collaborative work between remotely located people.

Jitter: Unevenness in the transmission of data, an important problem for streaming video.

Media Search Engine: A Web search engine designed to find media (usually images) on the Web.

Metadata: Information describing another data object such as its size or caption.

Multimedia: Data that include images, audio, video, or software.

Quality of Service: The quality of data transmission across a computer network as a composite of several factors.

Streaming: Video or audio sent in real time from a source, thereby reducing storage needs.

This work was previously published in Encyclopedia of Digital Government, edited by A. Anttiroiko & M. Malkia, pp.382-386,

Chapter 1.3
Core Principals of Educational Multimedia

Geraldine Torrisi-Steele
Griffith University, Australia

INTRODUCTION

The notion of using technology for educational purposes is not new. In fact, it can be traced back to the early 1900s during which school museums were used to distribute portable exhibits. This was the beginning of the visual education movement that persisted throughout the 1930s, as advances in technology such as radio and sound motion pictures continued. The training needs of World War II stimulated serious growth in the audiovisual instruction movement. Instructional television arrived in the 1950s but had little impact, due mainly to the expense of installing and maintaining systems. The advent of computers in the 1950s laid the foundation for CAI (computer assisted instruction) through the 1960s and 1970s. However, it wasn't until the 1980s that computers began to make a major impact on education (Reiser, 2001). Early applications of computer resources included the use of primitive simulation. These early simulations had little graphic capabilities and did little to enhance the learning experience (Munro, 2000).

Since the 1990s, there have been rapid advances in computer technologies in the area of multimedia production tools, delivery, and storage devices. Throughout the 1990s, numerous CD-ROM educational multimedia software was produced and was used in educational settings. More recently, the advent of the World Wide Web (WWW) and associated information and communications technologies (ICT) has opened a vast array of possibilities for the use of multimedia technologies to enrich the learning environment. Today, educational institutions are investing considerable effort and money into the use of multimedia. The use of multimedia technologies in educational institutions is seen as necessary for keeping education relevant to the 21st century (Selwyn & Gordard, 2003).

The term *multimedia* as used in this article refers to any technologies that make possible "the entirely digital delivery of content presented by using an integrated combination of audio, video, images (two-dimensional, three-dimensional) and text," along with the capacity to support user interaction (Torrisi-Steele, 2004,

p. 24). Multimedia encompasses related communications technologies such as e-mail, chat, video-conferencing, and so forth. "The concept of interaction may be conceptualised as occurring along two dimensions: the capacity of the system to allow individual to control the pace of presentation and to make choices about which pathways are followed to move through the content; and the ability of the system to accept input from the user and provide appropriate feedback to that input.… Multimedia may be delivered on computer via CD-ROM, DVD, via the internet or on other devices such as mobile phones and personal digital assistants or any digital device capable of supporting interactive and integrated delivery of digital audio, video, image and text data" (Torrisi-Steele, 2004, p. 24).

The fundamental belief underlying this article is that the goal of implementing multimedia into educational contexts is to exploit the attributes of multimedia technologies in order to support deeper, more meaningful learner-centered learning. Furthermore, if multimedia is integrated effectively into educational contexts, then teaching and learning practice must necessarily be transformed (Torrisi-Steele, 2004). It is intended that this article will serve as a useful starting point for educators beginning to use multimedia. This article attempts to provide an overview of concepts related to the effective application of multimedia technologies to educational contexts. First, constructivist perspective is discussed as the accepted framework for the design of multimedia learning environments. Following this, the characteristics of constructivist multimedia learning environments are noted, and then some important professional development issues are highlighted.

THEORETICAL FOUNDATIONS FOR THE ROLE OF MULTIMEDIA IN EDUCATIONAL CONTEXTS

Traditionally, teaching practices have focused on knowledge acquisition, direct instruction, and the recall of facts and procedures. This approach suited the needs of a society needing "assembly line workers" (Reigeluth, 1999, p. 18). However, in today's knowledge-based society, there is a necessity to emphasize deeper learning that occurs through creative thinking, problem solving, analysis, and evaluation, rather than the simple recall of facts and procedures emphasized in more traditional approaches (Bates, 2000). The advent of multimedia technologies has been heralded by educators as having the capacity to facilitate the required shift away from traditional teaching practices in order to innovate and improve on traditional practices (LeFoe, 1998; Relan & Gillani, 1997). Theoretically, the shift away from traditional teaching practices is conceptualized as a shift from a teacher-centered instructivist perspective to a learner-centered constructivist perspective on teaching and learning.

The constructivist perspective is widely accepted as the framework for design of educational multimedia applications (Strommen, 1999). The constructivist perspective describes a "theory of development whereby learners build their own knowledge by constructing mental models, or schemas, based on their own experiences" (Tse-Kian, 2003, p. 295). The constructivist view embodies notions that are in direct opposition to the traditional instructivist teaching methods that have been used in educational institutions for decades (see Table 1).

Expanding on Table 1, learning environments designed on constructivist principles tend to result in open-ended learning environments in which:

- Learners have different preferences of learning styles, cognitive abilities, and prior knowledge; they construct knowledge

Table 1. Key principles of the constructivist view of teaching and learning vs. key principles of the instructivist view of teaching and learning

CONSTRUCTIVIST	INSTRUCTIVIST
• learner-centered perspective: the learner is the focus of the learning environment – learners as individuals	• teacher-centered perspective: the teacher is focus of the learning environment- group learning
• encourages student independence in learning	• encourages student dependence on teacher
• teacher as facilitator that acts as a guide	• teacher as instructor
• learner and facilitator engage in a collaborative learning experience	• teacher in control of learning and in position of power over learner
• learners actively constructing knowledge in their own individual manner	• learners passively acquiring knowledge from the instructor
• Process of knowledge acquisition is important - how are learners interacting with the learning environment?	• acquisition of content and factual knowledge is key objective of learning episode
• curriculum design as development of knowledge spaces which allow active exploration by the learner	• curriculum design as goal oriented, strictly structured and ordered knowledge transmission
• Higher order thinking skills emphasized, creative thinking, problem solving, evaluation, synthesis	• behavioral objectives focusing on recall of facts and procedures, surface learning
• Open-ended learning environments (**OELE**)	• **directed instruction**

in individual ways by choosing their own pathways. Learning is affected by its contexts as well as the beliefs and attitudes of the learner;

- Optimal learning occurs when learners are active learners (e.g., learn by doing and learn by discovery;
- Learning is a process of construction whereby learners build knowledge through a process of scaffolding. Scaffolding is the process whereby learners link new knowledge with existing knowledge;
- Knowledge construction is facilitated through authentic problem-solving experiences;

- The process of learning is just as important as learning outcomes. Learners are encouraged to "articulate what they are doing in the environment and reasons for their actions" (Jonassen, 1999, p. 217).

Multimedia, by virtue of its capacity f or interactivity, media integration, and communication, can be easily implemented as a tool for information gathering, communication, and knowledge construction. Multimedia lends itself well to the "creation and maintenance of learning environments which scaffold the personal and social construction of knowledge" (Richards & Nason, 1999). It is worth noting that the interactivity attribute of multimedia is considered extremely important from a constructivist perspective. Interactivity in terms of navigation allows learners to take responsibility for the pathways they follow in following learning goals. This supports the constructivist principles of personal construction of knowledge, learning by discovery, and emphasis on process and learner control. Interactivity in terms of feedback to user input into the system (e.g., responses to quizzes, etc.) allows for guided support of the learner. This also is congruent with constructivist principles of instruction as facilitation and also consistent with the notion of scaffolding, whereby learners are encouraged to link new to existing knowledge.

Using the constructivist views as a foundation, the key potentials of multimedia to facilitate constructivist learning are summarized by Kramer and Schmidt (2001) as:

- Cognitive flexibility through different accesses for the same topic;
- Multi-modal presentations to assist understanding, especially for learners with differing learning styles;
- "Flexible navigation" to allow learners to explore "networked information at their own pace" and to provide rigid guidance, if required;
- "Interaction facilities provide learners with opportunities for experimentation, context-dependent feedback, and constructive problem solving";
- Asynchronous and synchronous communication and collaboration facilities to bridge geographical distances; and
- Virtual laboratories and environments can offer near authentic situations for experimentation and problem solving.

THE EFFECTIVE IMPLEMENTATION OF MULTIMEDIA IN EDUCATIONAL CONTEXTS

Instructional Design Principles

Founded on constructivist principles, Savery and Duffy (1996) propose eight constructivist principles useful for guiding the instructional design of multimedia learning environments:

- Anchor all learning activities to a larger task or problem.
- Support learning in developing ownership for the overall problem or task.
- Design an authentic task.
- Design the tasks and learning environment to reflect the complexity of the environment that students should be able to function in at the end of learning.
- Give the learner ownership of the process to develop a solution.
- Design the learning environment to support and challenge the learner's thinking.
- Encourage testing ideas against alternative views and contexts.
- Provide opportunity for and support reflection on both the content learned and the process itself.

Along similar lines, Jonassen (1994) summarizes the basic tenets of the constructivist-guided

instructional design models to develop learning environments that:

- Provide multiple representations of reality;
- Represent the natural complexity of the real world;
- Focus on knowledge construction, not reproduction;
- Present authentic tasks (contextualizing rather than abstracting instruction);
- Provide real-world, case-based learning environments rather than pre-determined instructional sequences;
- Foster reflective practice;
- Enable context-dependent and content-dependent knowledge construction; and support collaborative construction of knowledge through social negotiation, not competition among learners for recognition.

Professional Development Issues

While multimedia is perceived as having the potential to reshape teaching practice, oftentimes the attributes of multimedia technologies are not exploited effectively in order to maximize and create new learning opportunities, resulting in little impact on the learning environment. At the crux of this issue is the failure of educators to effectively integrate the multimedia technologies into the learning context.

[S]imply thinking up clever ways to use computers in traditional courses [relegates] technology to a secondary, supplemental role that fails to capitalise on its most potent strengths. (Strommen, 1999, p. 2)

The use of information technology has the potential to radically change what happens in higher education...every tutor who uses it in more than a superficial way will need to re-examine his or her approach to teaching and learning and adopt new strategies. (Tearle, Dillon, & Davis, 1999, p. 10)

Two key principles should underlie professional development efforts aimed at facilitating the effective integration of technology in such a way so as to produce positive innovative changes in practice:

Principle 1: Transformation in Practice as an Evolutionary Process

Transformation of practice through the integration of multimedia is a process occurring over time that is best conceptualized perhaps by the continuum of stages of instructional evolution presented by Sandholtz, Ringstaff, and Dwyer (1997):

- **Stage one:** Entry point for technology use where there is an awareness of possibilities, but the technology does not significantly impact on practice.
- **Stage two:** Adaptation stage where there is some evidence of integrating technology into existing practice
- **Stage three:** Transformation stage where the technology is a catalyst for significant changes in practice.

The idea of progressive technology adoption is supported by others. For example, Goddard (2002) recognizes five stages of progression:

- **Knowledge stage:** Awareness of technology existence.
- **Persuasion stage:** Technology as support for traditional productivity rather than curriculum related.
- **Decision stage:** Acceptance or rejection of technology for curriculum use (acceptance leading to supplemental uses).
- **Implementation stage:** Recognition that technology can help achieve some curriculum goals.

- **Confirmation stage:** Use of technology leads to redefinition of learning environment—true integration leading to change.

The recognition that technology integration is an evolutionary process precipitates the second key principle that should underlie professional development programs—reflective practice.

Principle 2: Transformation is Necessarily Fueled by Reflective Practice

A lack of reflection often leads to perpetuation of traditional teaching methods that may be inappropriate and thus fail to bring about "high quality student learning" (Ballantyne, Bain & Packer, 1999, p. 237). It is important that professional development programs focus on sustained reflection on practice from the beginning of endeavors in multimedia materials development through completion stages, followed by debriefing and further reflection feedback into a cycle of continuous evolution of thought and practice. The need for educators to reflect on their practice in order to facilitate effective and transformative integration of multimedia technologies cannot be understated.

In addition to these two principles, the following considerations for professional development programs, arising from the authors' investigation into the training needs for educators developing multimedia materials, are also important:

- The knowledge-delivery view of online technologies must be challenged, as it merely replicates teacher-centered models of knowledge transmission and has little value in reshaping practice;
- Empathising with and addressing concerns that arise from educators' attempts at innovation through technology;
- Equipping educators with knowledge about the potential of the new technologies (i.e., online) must occur within the context of the total curriculum rather than in isolation of the academic's curriculum needs;
- Fostering a team-orientated, collaborative, and supportive approach to online materials production;
- Providing opportunities for developing basic computer competencies necessary for developing confidence in using technology as a normal part of teaching activities.

LOOKING TO THE FUTURE

Undeniably, rapid changes in technologies available for implementation in learning contexts will persist. There is no doubt that emerging technologies will offer a greater array of possibilities for enhancing learning. Simply implementing new technologies in ways that replicate traditional teaching strategies is counterproductive. Thus, there is an urgent and continuing need for ongoing research into how to best exploit the attributes of emerging technologies to further enhance the quality of teaching and learning environments so as to facilitate development of lifelong learners, who are adequately equipped to participate in society.

CONCLUSION

This article has reviewed core principles of the constructivist view of learning, the accepted framework for guiding the design of technology-based learning environments. Special note was made of the importance of interactivity to support constructivist principles. Design guidelines based on constructivist principles also were noted. Finally, the importance of professional development for educators that focuses on reflective practice and evolutionary approach to practice transformation was discussed. In implementing future technologies in educational contexts, the

goal must remain to improve the quality of teaching and learning.

REFERENCES

Ballantyne, R., Bain, J.D., & Packer, J. (1999). Researching university teaching in Australia: Themes and issues in academics' reflections. *Studies in Higher Education, 24*(2), 237-257.

Bates, A.W. (2000). *Managing technological change*. San Francisco: Jossey-Bass.

Goddard, M. (2002). What do we do with these computers? Reflections on technology in the classroom. *Journal of Research on Technology in Education, 35*(1), 19-26.

Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods and models. In C. Reigeluth (Ed.), *Instructional-design theories and models* (pp. 115-140). Hillsdale, NJ: Erlbaum.

Jonassen, D.H. (1994). Thinking technology: Toward a constructivist design model. *Educational Technology, Research and Development, 34*(4), 34-37.

Jonassen, D.H. (1999). Designing constructivist learning environments. In C. Reigeluth (Ed.), *Instructional-design theories and models* (pp. 215-239). Hillsdale, NJ: Erlbaum.

Kramer, B.J., & Schmidt, H. (2001). Components and tools for on-line education. *European Journal of Education*, 36(2), 195-222.

Lefoe, G. (1998). *Creating constructivist learning environments on the Web: The challenge of higher education*. Retrieved August 10, 2004, from *http://www.ascilite.org.au/conferences/wollongong98/ascpapers98.html*

Munro, R. (2000). Exploring and explaining the past: ICT and history. *Educational Media International, 37*(4), 251-256.

Reigeluth, C. (1999). What is instructional-design theory and how is it changing? In C. Reigeluth (Ed.), *Instructional-design theories and models* (pp. 5-29). Hillsdale, NJ: Erlbaum.

Reiser, R.A. (2001). A history of instructional design and technology: Part I: A history of instructional media. *Educational Technology, Research and Development, 49*(1), 53-75.

Relan, A., & Gillani, B. (1997). Web-based instruction and the traditional classroom: Similarities and differences. In B.H. Khan (Ed.), *Web-based instruction* (pp. 41-46). Englewood Cliffs, NJ: Educational Technology Publications.

Richards, C., & Nason, R. (1999). Prerequisite principles for integrating (not just tacking-on) new technologies in the curricula of tertiary education large classes. In J. Winn (Ed.) *ASCILITE '99 Responding to diversity conference proceedings.* Brisbane: QUT. Retrieved March 9, 2005 from *http://www.ascilite.org.au/conferences/brisbane99/papers/papers.htm*

Sandholtz, J., Ringstaff, C., & Dwyer, D. (1997). *Teaching with technology.* New York: Teachers College Press.

Savery J.R. & Duffy T.M. (1996). An instructional model and its constructivist framework. In B Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design.* Englewood Cliffs, NJ: Educational Technology Publications.

Selwyn, N., & Gorard, S. (2003). Reality bytes: Examining the rhetoric of widening educational participation via ICT. *British Journal of Educational Technology, 34*(2), 169-181.

Strommen, D. (1999). *Constructivism, technology, and the future of classroom learning.* Retrieved September 27, 1999, from *http://www.ilt. columbia.edu/ilt/papers/construct.html*

Tearle, P., Dillon, P., & Davis, N. (1999). Use of information technology by English university

teachers. Developments and trends at the time of the national inquiry into higher education. *Journal of Further and Higher Education, 23*(1), 5-15.

Torrisi, G., & Davis, G. (2000). Online learning as a catalyst for reshaping practice—The experiences of some academics developing online materials. *International Journal of Academic Development, 5*(2), 166-176.

Torrisi-Steele, G. (2004). Toward effective use of multimedia technologies in education In S. Mishra & R.C. Sharma (Eds.), *Interactive multimedia in education and training* (pp. 25-46). Hershey, PA: Idea Group Publishing.

Tse-Kian, K.N. (2003). Using multimedia in a constructivist learning environment in the Malaysian classroom. *Australian Journal of Educational Technology, 19*(3), 293-310.

KEY TERMS

Active Learning: A key concept within the constructivist perspective on learning that perceives learners as mentally active in seeking to make meaning.

Constructivist Perspective: A perspective on learning that places emphasis on learners as building their own internal and individual representation of knowledge.

Directed Instruction: A learning environment characterized by directed instruction is one in which the emphasis is on "external engineering" (by the teacher) of "what is to be learned" as well as strategies for "how it will be learned" (Hannafin, Land & Oliver, 1999, p. 122).

Instructivist Perspective: A perspective on learning that places emphasis on the teacher in the role of an instructor that is in control of what is to be learned and how it is to be learned. The learner is the passive recipient of knowledge. Often referred to as teacher-centered learning environment.

Interactivity: The ability of a multimedia system to respond to user input. The interactivity element of multimedia is considered of central importance from the point of view that it facilitates the active knowledge construction by enabling learners to make decisions about pathways they will follow through content.

Multimedia: The entirely digital delivery of content presented by using an integrated combination of audio, video, images (two-dimensional, three-dimensional) and text, along with the capacity to support user interaction (Torrisi-Steele, 2004).

OELE: Multimedia learning environments based on constructivist principles tend to be open-ended learning environments (OELEs). OELEs are open-ended in that they allow the individual learner some degree of control in establishing learning goals and/or pathways chosen to achieve learning.

Reflective Practice: Refers to the notion that educators need to think continuously about and evaluate the effectiveness of the strategies and learning environment designs they are using.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 130-136, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.4
Multimedia Instruction

Lorna Uden
Staffordshire University, UK

INTRODUCTION

Multimedia technology is increasingly being used as a vehicle to deliver instruction.

The terms "hypermedia" and "multimedia" are often used interchangeably. However, a distinction is sometimes made: Not all multimedia applications are necessarily hypermedia. A network representation of information is one of the defining characteristics of hypermedia. An instance of hypermedia consists of pieces of information connected in an arbitrary manner to form a network of references (Begoray, 1990). In this chapter, the terms will be used synonymously. There are many benefits for using multimedia for instruction.

Studies have shown that computer-based multimedia can help people learn more information better than traditional classroom lectures (Bagui, 1998). Several factors have been attributed to the success of multimedia in helping people to learn. First, there is a parallel between multimedia and the 'natural' way people learn, as explained by the Information Processing Theory (Gagné, Briggs & Wager, 1988). The similarities between the structure of multimedia and the information processing theory account for a large part of the success of learning with multimedia.

Second, information in computer-based multimedia is presented in a non-linear, hypermedia format. The nature of hypermedia allows learners to view things from different perspectives. Third, computer-based multimedia is more interactive than traditional classroom lectures. Interacting appears to have a strong positive effect on learning (Najjar, 1996). Fourth, another feature of multimedia-based learning is that of flexibility. There is empirical evidence (Najjar, 1996) that interactive multimedia information helps people learn.

A large number of presentation guidelines have been reported for educational multimedia (Kozma, 1991; Park & Hannafin, 1994) that advise on selecting certain media for different types of content and learning goals. Investigation by Schaife and Rogers (1996) revealed that many of these products exhibit poor usability and are ineffective in learning. Multimedia design is currently created by intuition (Sutcliffe, 1997). Given the complexity of multimedia interaction,

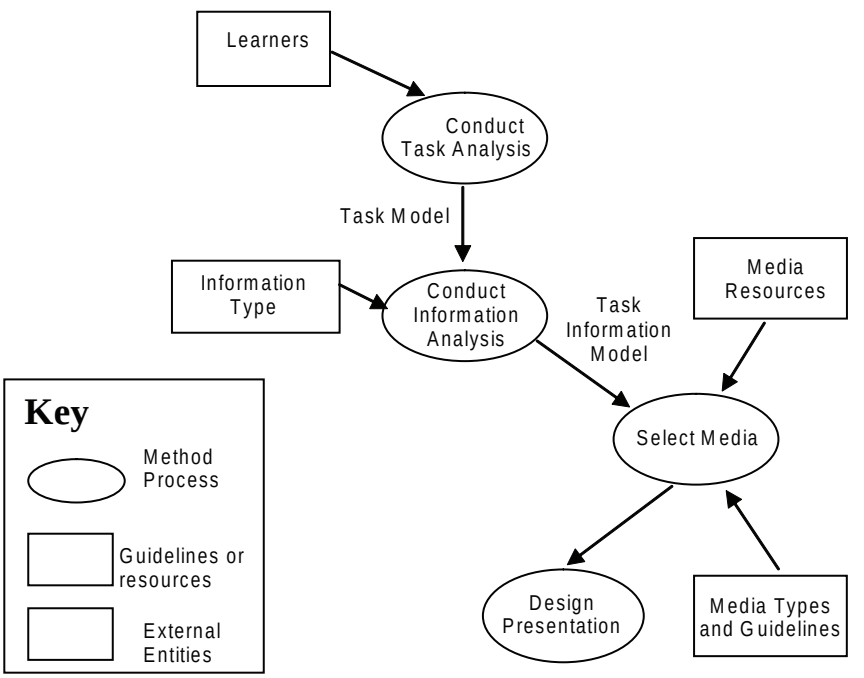
it is unlikely that the craft-style approach will produce effective interfaces. A methodical approach to multimedia interface design is needed. Guidelines are required to cover selection of media resources for representing different types of information and presentation design. These guidelines must address the key issues of selective attention, persistence of information, concurrency and preventing information overload. Multimedia provides designers with many opportunities to increase the richness of the learner interface, but with richness comes the penalty that interfaces can become overcrowded with too much information. Using multimedia does not ensure that information is conveyed in a comprehensive manner. Careful design is required to ensure that the medium matches the message and that important information is delivered effectively. To date, there are few methods available that give detailed guidelines to help designers choose the most appropriate medium based on the information types required. Subsequent sections of this article describe a method known as Multimedia Instructional Design Method (MIDM), based on the Sutcliffe and Faraday method (1994) that can be used to help novice designers to develop multimedia instruction.

MULTIMEDIA INSTRUCTION DESIGN METHOD

The method developed is based on the work of Sutcliffe and Faraday (1994). It consists of four main stages: task analysis, information analysis, media selection and presentation, as shown in Figure 1.

The method consists of design principles based on cognitive psychology, media selection guidelines that utilise definition of information and media types, validation guidelines for media combination and techniques for directing the learners' attention in multimedia sequences.

Figure 1. Overview of the method



The first step is the creation of a task model incorporating specification of the content information requirements. A resource model describing the information media available to the designer then follows this. The method advises on selecting appropriate media for the information needs and scripting a coherent presentation for a task context.

The next design step is to direct the learner's attention to extract the required information from a given presentation and focus on the correct level of detail. This forces designers to be aware of the cognitive issues underlying a multimedia presentation, such as selective attention, persistence of information, concurrency and limited cognitive resources, such as working memory.

Step 1: Conduct Task Analysis

The method starts with a standard task analysis using one of the instructional task analysis methods. Both hierarchical task analysis and information-passing task analysis methods can be used. In addition, a learner analysis would also be conducted. The task analysis would have produced a hierarchy of goals composed of sub-goals, which in turn contain procedures, actions and objects.

Step 2: Conduct Information Analysis

The task analysis does not explicitly state what information is required for each task step, so an information analysis is needed. The main objective of information analysis is to specify what type of information is required during a task, called the task information model. To form the task information model, the initial goal hierarchy from the task analysis model is elaborated on by attaching information types that specify the content to be communicated to the learner. The resulting model should allow the designer to answer the question, "What information content does the learner need for this task sub-goal or input/output interaction?" A set of amodal information types is required to characterise lesson needs. Information types are used to specify the message to be delivered in a multimedia application and are operated on by mapping rules that select the appropriate media types. The information types are similar to those found in many tasks or data models (e.g., actions,

Table 1. Type of information

Abstract information	Facts or objects that do not have a physical existence; e.g., human knowledge, facts, concepts, plans.
Causation	Description of the cause and effect of an event, including a sequence of events that describe causation; e.g., heat causing water to boil, behaviour of an algorithm that results in a desired goal.
Composition	The aggregation or assembly of an object, whole-part relationships; e.g., components of a car engine, parts of a computer processor, memory, motherboard.
Description	Facts that describe an object, entity or agent; e.g., red apples, texture of stone.
States	Descriptions that remain constant for part of the world, objects or agent world, during a period of time; e.g., a person is sleeping.
Physical	Objects or agents that have a physical existence; e.g., chair, table.
Visio-spatial	Visual attributes of objects, structures, pathways, spatial distribution, location, size, shape; e.g., layout of furniture in a room, direction to bus station, shape of a mountain.

objects, procedures). Task actions may require operational information (the nature of the action to be performed), temporal information (the time course of the action to be performed) or spatial information (the physical nature of the action). Task objects may require descriptive information (detail of the object) or spatial information (the location of the object). Information types are amodal or conceptual descriptions of information components that elaborate the content definition. Table 1 describes some of the information types that can be used.

Information types are used to help refine descriptions of the necessary content. The motivation is to provide informal categories that help assess what type of information is required to support user and system tasks. Information can be broadly divided into static data about objects and dynamic data describing actions, events and changes in the environment.

A media resource analysis is also carried out in this stage. This section describes media resources available to the application for presentation. Media resources considered here are linguistic (text and speech), still image (photographs, drawings) and moving image (animated diagrams, video). Classification is based on the psychological properties of the representation rather than the physical nature of the medium (e.g., digital or analogue encoding in video). Table 2 lists the media type definitions and examples. These definitions are combined to describe any specific medium, so speech is classed as an audio, linguistic medium, while a cartoon is classified as a non-realistic (designed) moving image. Each media resource will contain one or more information types: temporal information, spatial information, operational information or descriptive information.

Having defined the resources available for presentation, the information task model and the information requirements for the task, we now have to decide which media types to use for the task's information needs. To do this, the task model is first elaborated on by attaching dialogue acts to specify the desired effect of presenting the information. Dialogue acts are used in explanation systems (Mann & Thompson, 1988). These are added to the task model to designate the

Table 2. Media type definitions and examples

Media Type	Description	Example
Non-realistic	content created by human action rather than being captured from the real world.	Diagrams, graphs, cartoons.
Realistic	content perceived by users to have been captured form the natural world rather than being explicitly designed.	Natural sounds, photographic images, film showing people and natural scenes.
Audio	any medium received by the audio channel (hearing), sounds	Dog barking, music, traffic noise, speech.
Linguistic	text, spoken language and symbols interpreted in linguistic terms.	Alphanumeric text, symbols and signs.
Moving image	visual media delivered at a continuous rate.	Video, film, animated diagrams and simulations.
Still image	visual media that are not presented continuously.	Photographs, drawings, graphs.

Table 3. Dialogue acts and their communicative effects

Act Type	Act Name	Communicative Effect
Subject-informative	Enable	Communicate actions to achieve a task sub-goal.
	Result	Give information about the outcome of a task sub-goal.
	Cause	Give information concerning the causality of a task sub-goal.
	Inform	Display information as-is.
Subject-organising	Sequence	Specify a succession of linked steps.
	Summary	Provide overview of task sub-goal(s).
	Condition	A particular task sub-goal is a pre-condition.
Presentational	Locate	Draw attention to an information type.
	Foreground	Give further detail of an information type.
	Background	Give content information.
	Emphasise	Make an information type prominent.

communicative effect to be achieved, answering the questions: "What information does the user need for the goal?" and "How should the user's attention be drawn to important information?" These questions can be answered by splitting the dialogue acts into subject and presentation based acts (see Table 3).

Subject-based acts elaborate on information types according to the procedure and action-level detail in the task model. Subject acts are subdivided into subject-informative acts (which define information needs according to stages of task procedure) and subject-organising acts (which control the sequence of presentation and hence organise the subject-informative acts).

Presentational acts are used on media resources to draw the user's attention to particular information type(s), thereby supporting the subject act. Dialogue acts are linked to the task model by "walkthrough," asking, "What information or explanation is required at this step?"

At the start of each task procedure, describe its order (sequence), the goal (summary) and any preconditions upon it being performed (conditions). Information may be necessary as input for task actions (enable), followed by display of results of action (result). The user's needs could be to "Foreground" the information type (e.g., draw attention to a property of object associated with a particular task action) or to present an information type at a higher, more general level such as may be necessary at the start of a task (summary), or to draw attention explicitly to important information (emphasise).

Step 3: Select Media

The incorporation of media in a multimedia instructional application requires design principles based on modality theory. Modality is the medium used; hence, it is the mode of representation to present the required message. There are thousands of modalities, both input and output, that can be incorporated into interface designs (Bernsen, 1994). Selecting an optimal unimodal modality from this vast array of alternatives is difficult, due to each modality having a set of information representation characteristics, making it good for

the representation of certain information types and bad for others. The combination of two or more of these modalities exacerbates the problem as, when several modalities (both input and/or output) are involved, media interference needs to be taken into consideration.

According to Bernsen (1994), Modality Theory addresses the following general information-mapping problem: "Given any particular set of information which needs to be exchanged between the user and system during task performance in context, identify the input/output modalities which constitute an optimal solution to the representation and exchange of the information" (p. 348). Current multimedia research attempts to address modality design by creating methods that remove the ad hoc nature of solutions by providing theoretical frameworks for developers to follow. The problem these methods must solve, if viewed in the most basic terms, can be regarded as the information-mapping problem, which requires that a mapping exist between task requirements and a set of usable modalities.

Several issues need to be addressed when designing multimedia applications. When selecting a medium for use, we should select one that is specifically suited/adapted to represent the information requirements of the concept; that is, to describe the appearance of a person, a photograph should be used in preference to a text extract, as it provides a clearer description. This involves the use of theory on media usage, because the aim is to select a medium that best addresses the information requirements of the concept to be represented. The way to do this is to use media selection rules derived from modality theory. Another issue to be considered is how can optimal media or combinations be selected? These issues highlight the need for theoretical methods. Combining methods with media theory enables us to derive principles that can guide us toward making optimal media and combination decisions.

Media Selection Guidelines

When different media resources are available, we need to choose the medium to deliver the information needed. Task and user characteristics influence media choice. For instance, verbal media are more appropriate to language-based and logical reasoning tasks. Conversely, visual media are suitable for spatial tasks, including moving, positioning and orienting objects. Selection guidelines are based on the information types required for the subject dialogue act and task step type. The information types used with the media selection guidelines are shown in Table 4.

The guidelines may be used in multiple passes. For example, when a procedure for explaining a physical task is required, we can first call realistic image media, then a series of images and text. Guidelines that differentiate physical from abstract information are used first, followed by other guidelines.

Having designed the presentation, a set of validation guidelines is applied to the presentation design. The guidelines are derived from psychological and instructional design literature. The following guidelines are adopted from Sutcliffe and Faraday (1994):

1. If visual and verbal modalities are used concurrently, ensure congruent presentation by checking that information on each modality is semantically related.
2. Use text or still images for key messages.
3. Use verbal channel for warning.
4. Multiple attention-gaining devices should be avoided within a single task sub-goal presentation.
5. Do not present different subject matter on separate channels.
6. Use only visual-verbal channels concurrently.
7. Present only one animation or changing still image resource at a time.

Table 4. Overview of media selection preferences and examples

Information types	Preferred Media Selection	Example
Physical	still or moving image	building diagram
Abstract	text or speech	explain sales policy
Descriptive	text or speech	chemical properties
Visio-spatial	realistic media – photograph	person's face
Value	text/tables/numeric lists	pressure reading
Relationships (values)	design images – graphs – charts	histogram of rainfall per month
Procedural	images, text	evacuation instruction
Discrete action	still image	make of coffee
Continuous action m	oving image	manoeuvres while skiing
Events	sound, speech	fire alarm
States	still image, text	photo of weather conditions
Composition	still image, moving image	exploded part diagrams
Causal	moving image, text, speech	video of rainstorm causing flash flood

8. **Beware:** Visual media tend to dominate over verbal—place important messages in visual channel in concurrent visual-verbal presentations.
9. If several information types need to be semantically integrated, then use a common modality, as associations are formed more effectively within than between single semantic systems.

Step 4: Design the Presentation

Sutcliffe (1999) states that presentation design is primarily concerned with visual media, as the user's viewing sequence is unpredictable. The design should make important information salient in both speech and sound. It is important that a designer in multimedia links the thread of a message across several media. Presentation techniques are used to help direct the learner's attention to important information and to specify the desired order of reading or viewing. While the information model defines the high-level presentation order, presentation bar charts are used to plan the sequence and duration of media delivery.

First, the information types are ordered in a "first cut" sequence. The selected media are then added to the bar chart to show which media stream will be played over time. Decisions on timing depend on the content of the media resource; for example, length of a video clip, frame display and rate.

It is important to add focus-control actions to the first-cut presentation script to either make specific information within a medium more salient or to draw the learners' attention to message links between media. The need for focus shifts between information components is identified. Attention marker techniques are selected to implement the desired effect. Things to consider here are:

1. Plan the overall thematic thread of the message through the presentation or dialogue.
2. Draw the learners' attention to important information.
3. Establish a clear reading/viewing sequence.
4. Provide clear links when the theme crosses from one medium to another.

Direct Learner Attention Techniques

When the message is important and cross-referencing is critical, it is important to draw the learners' attention to both the source and destination medium. "Contact point" is used to describe a reference from one medium to another (Sutcliffe, 1999). There are two types of contact point: direct

and indirect. With direct contact points, attention-directing effects are implemented in both source and destination media. Conversely, with indirect contact points, an attention-directing effect is implemented only in the source medium.

Guidelines for Contact Point Uses

1. **Direct contact points for key thematic links:** A direct contact point should be used if the connection between information in two different media is important. For example, speech is used to direct the learners to the object in the image while highlighting the object being spoken about: "Look at the map"; the location of the laboratory is highlighted; or a text caption is revealed with an arrow pointing to the laboratory.
2. **Direct contact points for linked components:** Direct contact points should be used if components in both source and destination media are important and have to be perceived. For example: "Locate fire team in the diagram" (speech track), the team location is highlighted (with arrow).
3. **Indirect contact points:** Indirect contact points are used when the connection between information in two media is necessary, but perception of the destination component is less important. For example: "Look at the diagram," speaking about the object while displaying the image.

This concludes the method stages, which have now produced a detailed and thematically integrated presentation design. The guidelines can be applied either to the specification bar chart before implementation or interactively during a cycle of prototype implementation, evaluation and critiquing.

FUTURE TRENDS

Multimedia development is currently at the stage that software development was 30 years ago. Most applications are developed using an ad hoc approach. There is little understanding of development methodologies, measurement and evaluation technologies, development processes, and application quality and project management. This current approach to developing multimedia is, in many cases, failing to deliver applications that have acceptance quality, especially in terms of information access and usability. According to researchers (Lowe & Hall, 1999), this failure is largely due to a lack of process. As multimedia applications grow in scope and complexity, we need an evolution (or revolution) to occur, similar to that which occurred in software development. Just as the focus in software development shifted from programming to process, the focus with multimedia must shift from the use of specific authoring tools in handcrafting applications to broader process issues that support the development of high-quality, large-scale applications. This includes aspects such as framework, tools and techniques, validation methods and metrics.

Although multimedia development tools are important, as with software tools, they must be used appropriately within an overall development process that gives them a suitable context. Small applications can still be readily handcrafted, but for large applications, and especially those which will evolve over time, we need to adopt a more formal and thorough approach; that is, a "multimedia engineering" approach. Multimedia engineering is the employment of a systematic, disciplined, quantifiable approach to the development, operation and maintenance of multimedia applications. Applying an engineering approach to multimedia development underlines two primary emphases: first, that multimedia development is a process. This process includes more than just media manipulation and presentation creation. It includes analysis of needs, design management,

metrics and maintenance. The second emphasis is the handling and management of information to achieve some desired goal. Hypermedia engineering is the combination of these two emphases—the use of suitable processes in creating multimedia applications effective in managing and utilising information.

CONCLUSION

The design and development of multimedia instructional applications is not trivial. To produce effective multimedia applications, it is necessary that guidelines used must be based on principles of cognitive psychology. The method proposed is, we believe, a first comprehensive multimedia presentation design method. Although many guidelines have been proposed by researchers to direct designers in their multimedia development, none of these researchers have integrated their guidelines into a design method. One criticism that may be levelled against all of these guidelines is that they produce recommendations expert designers would have come up with anyway. This misses the point that methods *are* produced and introduced to help novice designers improve their performance and ensure that they achieve at least an adequate standard.

Evidence produced by Sutcliffe (1997, 1999) has shown that the proposed method did help improve multimedia applications. Our own experience in using the method to develop multimedia instructional applications has been very encouraging. The above method has been used successfully by more than 60 students from the multimedia degree course in our university to develop different multimedia applications. So far, the method has proved useful as a means of exploring the issues involved in multimedia design. The method provides a tool for thought about presentation issues concerning what information is required and when.

However, the guidelines within the method need further usability testing and evaluation of their effectiveness. Many of the guidelines will appear in the forthcoming multimedia user-interface design standard, ISOI 14915. The method, while making no claim to have solved the multimedia interface design problem, has explored the issues that must be addressed in multimedia interface design. The guidelines and the method can, however, be used as sound design advice to help novice designers develop multimedia instruction.

REFERENCES

Bagui, S. (1998). Reasons for increased learning using multimedia. *Journal of Educational Multimedia and Hypermedia, 7*(7), 3-18.

Bernsen, N.O. (1994). Foundations of Multimodal Representations: a taxonomy of representational modalities. *Interacting with Computers, 6*, 347-371.

Gagné, R.M., Briggs, L.J., & Wager, W.W. (1988). *Principles of instructional design* (3rd edition). London: Holt, Rinehart & Winston.

Kozma, R.B. (1991). Learning with Media. *Review of Educational Research, 61*(2), 179-211.

Lowe, D., & Hall, W. (1999). *Hypermedia and the Web: An Engineering Approach*. New York: Wiley & Son.

Mann, W.C., & Thompson, S.A. (1988). A Rhetorical Structure Theory: toward a functional theory of text organisation. *Text, 8*(3), 243-281.

Najjar, L.J. (1996). Multimedia Information and Learning. *Journal of Multimedia and Hypermedia, 5*(2), 129-150.

Park, I., & Hannafin, J. (1994). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology Research & Development, 4*(3), 63-85.

Schaife, M., & Rogers, Y. (1996). External Cognition: how do graphical representations work? *International Journal of Human Computer Studies, 45,* 185-213.

Sutcliffe, A.G. (1997). Task Related Information Analysis. *International Journal of Human Computer Studies, 47,* 223-255.

Sutcliffe, A.G. (1999). A design method for effective information delivery in multimedia presentation. *The New Review of Hypermedia and Multimedia*.

Sutcliffe, G., & Faraday, P.F. (1994). Designing Presentation in Multi Media Interfaces. In B. Adelson, S. Dumais & J. Olson (Eds.), *Proceedings of CHI-94* (pp. 92-98). New York: ACM Press.

KEY TERMS

Contact Point: Used to describe a reference from one medium to another.

Education Multimedia: The use of multimedia for designing educational software.

Information Type: The description of necessary content in an application.

Modality: The medium used, the mode of representation to present the required message.

Modality Theory: Given any particular set of information that needs to be exchanged between the users and system during task performance in context, identify the input/output modalities that constitute an optimal solution to representation and exchange of the information (Bernsen, 1994).

Multimedia: The use of different media such as text, graphics, audio and video, and others.

Multimedia Engineering: The employment of a systematic, disciplined, quantifiable approach to the development, operation and maintenance of multimedia applications.

This work was previously published in Encyclopedia of Distance Learning, Vol. 3, edited by C. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers & G.A. Berg, pp. 1317-1324, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.5
Multimedia Technologies in Education

Armando Cirrincione
SDA Bocconi School of Management, Italy

WHAT ARE MULTIMEDIA TECHNOLOGIES

MultiMedia Technologies (MMT) are all that kind of technological tools that make us able to transmit information in a very large meaning, transforming information into knowledge through stimulating the cognitive schemes of learners and leveraging the learning power of human senses. This transformation can acquire several different forms: from digitalized images to virtual reconstructions, from simple text to iper-texts that allow customized, fast, and cheap research within texts; from communications framework like the Web to tools that enhance all our sense, allowing complete educational experiences (Piacente, 2002b).

MMT are composed by two great conceptually different frameworks (Piacente, 2002a):

- **Technological supports, as hardware and software:** all kinds of technological tools such as mother boards, displays, videos, audio tools, databases, communications software and hardware, and so on;
- **Contents:** information and to knowledge transmitted with MMT tools. Information are simply data (such as visiting timetable of museum, cost of tickets, the name of the author of a picture), while knowledge comes from information *elaborated in order to get a goal*. For instance, a complex ipertext about a work of art, where much information is connected in a logical discourse, is knowledge. For the same reason, a virtual reconstruction comes from knowledge about the rebuilt facts.

It's relevant to underline that to some extent technological supports represent a condition and a limit for contents (Wallace, 1995). In other words, content could be expressed just through technological supports, and this means that content has to be made in order to fit for specific technological support and that the limits of a specific technological support are also the limits of its content. For instance the specific architecture of a database represents a limit within which contents have to be recorded and have to be traced. This is also evident thinking about content as a communicative

action: communication is strictly conditioned by the tool we are using.

Essentially, we can distinguish between two areas of application of MMT (Spencer, 2002) in education:

1. Inside the educational institution (schools, museums, libraries), with regard to all tools that foster the value of lessons or visiting during time they takes place. Here we mean "enhancing" as enhancing moments of learning for students or visitors: hypertexts, simulation, virtual cases, virtual reconstructions, active touch-screen, video, and audio tools;
2. In respect of outside the educational institution, this is the case of communication technologies such as Web, software for managing communities, chats, forums, newsgroups, for long-distance sharing materials, and so on. The power of these tools lies on the possibilities to interact and to cooperate in order to effectively create knowledge, since knowledge is a social construct (Nonaka & Konno, 1998; von Foester, 1984; von Glaserfeld, 1984).

Behind these different applications of MMT lies a common database, the heart of the multimedia system (Pearce, 1995). The contents of both applications are contained into the database, and so the way applications can use information recorded into database is strictly conditioned by the architecture of database itself.

DIFFERENT DIMENSIONS OF MMT IN TEACHING AND LEARNING

We can distinguish two broader framework for understanding contributions of MMT to teaching and learning.

The first pattern concerns the place of teaching; while in the past, learning generally required the simultaneous presence of teacher and students for interaction, now it is possible to teach long distance, thanks to MMT.

The second pattern refers to the way people learn; they can be passive or they can interact. The interaction fosters learning process and makes it possible to generate more knowledge in less time.

Teaching on Site and Distance Teaching

Talking about MMT applications in education requires to separate learning on-site and distance learning, although both are called e-learning (electronic learning). E-learning is a way of fostering learning activity using electronic tools based on multimedia technologies (Scardamaglia & Bereiter, 1993).

The first pattern generally uses MMT tools as a support to traditional classroom lessons; the use of videos, images, sounds, and so on can dramatically foster the retention of contents in student's minds (Bereiter, Scardamaglia, Cassels, & Hewitt, 1997).

The second pattern, distance teaching, requires MMT applications for a completely different environment, where students are more involved in managing their commitment. In other words, students in e-learning have to use MMT applications more independently than they are required to do during a lesson on site. Although this difference is not so clear among MMT applications in education, and it is possible to get e-learning tools built as they had to be used during on-site lessons and vice-versa, it is quite important to underline the main feature of e-learning not just as a distant learning but as a more independent and responsible learning (Collins, Brown, & Newman, 1995).

There are two types of distance e-learning: self-paced and leader-led. The first one is referred to the process students access computer-based (CBT) or Web-based (WBT) training materials

at their own pace. Learners select what they wish to learn and decide when they will learn it.

The second one, leader-led e-learning, involves an instructor and learners can access real-time materials (synchronous) via videoconferencing or audio or text messaging, or they can access delayed materials (asynchronous).

Both the cited types of distance learning use performance support tools (PST) that help students in performing a task or in self-evaluating.

Passive and Interactive Learning

The topic of MMT applications in an educational environment suggests distinguishing two general groups of applications referring to required students behaviour: passive or interactive. Passive tools are ones teachers use just to enhance the explanation power of their teaching: videos, sounds, pictures, graphics, and so on. In this case, students do not interact with MMT tools; that means MMT application current contents don't change according to the behaviour of students.

Interactive MMT tools change current contents according to the behaviour of students; students can chose to change contents according with their own interests and levels. Interactive MMT tools use the same pattern as the passive ones, such as videos, sounds, and texts, but they also allow the attainment of special information a single student requires, or they give answers just on demand. For instance, self-evaluation tools are interactive applications. Through interacting, students can foster the value of time they spent in learning, because they can use it more efficiently and effectively.

Interaction is one of the most powerful instruments for learning, since it makes possible active cooperation in order to build knowledge. Knowledge is always a social construct, a sense-making activity (Weick, 1995) that consists in giving meaning to experience. Common sense-making fosters knowledge building thanks to the richness of experiences and meanings people can exchange. Everyone can express his own meaning for an experience, and interacting this meaning can be elaborated and it can be changed until it becomes common knowledge. MMT help this process since they make possible interaction in less time and over long distance.

THE LEARNING PROCESS BEHIND E-LEARNING

Using MMT applications in education allows to foster learning process since there are several evidences that people learn more rapidly and deeply from words, images, animations, and sounds, than from words alone (Mayer, 1989; Mayer and Gallini, 1990). For instance, in the museum sector there is some evidence of the effectiveness of MMT devices too: Economou (1998) found firstly that people spend more time and learn more within a museum environment where there are MMT devices.

The second reason why MMT fosters learning derives from interaction they make possible. MMT allow building a common context of meaning, to socialize individual knowledge, to create a network of exchanges among teacher and learners. This kind of network is more effective when we consider situated knowledge, a kind of knowledge adults require that is quite related to problem-solving.

Children and adults have different pattern of learning, since adults are more autonomous in the learning activity and they also need to refer new knowledge to the old one they possess. E-learning technologies have developed a powerful method in order to respond more effectively and efficiently to the needs of children and adults: the "learning objects" (LO). Learning objects are single, discrete modules of educational contents with a certain goal and target. Every learning object is characterized by content and a teaching method that foster a certain learning tool: intellect, senses (sight, heard, and so on), fantasy,

analogy, metaphor, and so on. In this way, every learner (or every teacher, for children) can choose its own module of knowledge and the learning methods that fit better with his own level and characteristics.

As far as the reason why people learn more with MMT tools, it is useful to consider two different theories about learning: the *information delivery theory* and the *cognitive theory.* The first one stresses teaching as just a delivery of information and it looks at students as just recipients of information.

The second one, the cognitive theory, considers learning as a sense-making activity and teaching as an attempt to foster appropriate cognitive processing in the learner. According to this theory, instructors have to enable and encourage students to actively process information: an important part of active processing is to construct pictorial and verbal representations of the lesson's topics and to mentally connect them. Furthermore, archetypical cognitive processes are based on senses, that means: humans learn immediately with all five senses, elaborating stimuli that come from environment. MTT applications can be seen as virtual reproductions of environment stimuli, and this is another reason why MMT can dramatically fostering learning through leveraging senses.

CONTRIBUTIONS AND EFFECTIVENESS OF MMT IN EDUCATION

MMT allow transferring information with no time and space constraints (Fahy, 1995).

Space constraints refer to those obstacles that arise from costs of transferring from one place to another. For instance, looking at a specific exhibition of a museum, or a school lesson, required to travel to the town where it happens; participating to a specific meeting or lesson that takes place in a museum or a school required to be there; preparing an exhibition required to meet work group daily. MMT allows the transmission of information everywhere very quickly and cheaply, and this can limit the space-constraint; people can visit an exhibition stay at home, just browsing with a computer connected on internet. Scholars can participate to meeting and seminars just connecting to the specific web site of the museum. People who are organizing exhibitions can stay in touch with the Internet, sending to each other their daily work at zero cost.

Time constraint has several dimensions: it refers to the need to catch something just when it takes place. For instance, a lesson requires to be attended when it takes place, or a temporary exhibition requires to be visited during the days it's open and just for the period it will stay in. For the same reason, participating in a seminar needs to be there when it takes place.

But time constraint refers also to the limits people suffer in acquiring knowledge: people can pay attention during a visit just for a limited period of time, and this is a constraint for their capability of learning about what they're looking for during the visiting.

Another dimension of time constraint refers to the problem of rebuilding something that happened in the past; in the museum sector, it is the case of extemporary art (body art, environmental installations, and so on) or the case of an archaeological site, and so on.

MMT help to solve these kinds of problems (Crean, 2002; Dufresne-Tassé, 1995; Sayre, 2002) by making it possible:

- To attend school lessons on the Web, using videostreamer or cd-rom, allowing repetition of the lesson or just one difficult passage of the lesson (solving the problem of decreasing attention over time);
- To socialize the process of sense making, and so to socialized knowledge, creating networks of learners;
- To prepare the visit through a virtual visit to the Web site: this option allows knowing ex-ante what we are going to visit, and

doing so, allows selection of a route more quickly and simply than a printed catalogue. In fact, thanks to iper-text technologies, people can obtain lot of information about a picture just when they want and just as they like. So MMT make it possible to organize information and knowledge about heritage into databases in order to customize the way of approaching cultural products. Recently the Minneapolis Institute of Art has started a new project on Web, projected by its Multimedia department, that allow consumers to get all kind of information to plan a deep organized visit;

- To cheaply create different routes for different kind of visitors (adults, children, researcher, academics, and so on); embodying these routes into high tech tools (PCpalm, LapTop) is cheaper than offering expensive and not so effective guided tours.
- To re-create and record on digital supports something that happened in the past and cannot be renewed. For instance the virtual re-creation of an archaeological site, or the recording of an extemporary performance (so diffuse in contemporary art).

For all the above reasons, MMT enormously reduces time and space constraints, therefore stretching and changing the way of teaching and learning.

REFERENCES

Bereiter C., Scardamalia M., Cassels C., & Hewitt J. (1997). Postmodernism, knowledge building and elementary sciences, *The Elementary School Journal, 97*(4), 329-341.

Collins, A., Brown, J.S., & Newman S. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In Resnick, L.B. (Ed.), *Cognition and instructions: Issues and agendas*, Lawrence Erlbaum Associates.

Crean B. (2002). Audio-visual hardware. In B. Lord & G.D. Lord (Eds.), *The manual of museum exhibitions*, Altamira Press.

Dufresne-Tassé, C. (1995). Andragogy (adult education) in the museum: A critical analysis and new formulation. In E. Hooper-Greenhill (Ed.), *Museum, media, message*, London: Routledge.

Economou, M. (1998). The evaluation of museum multimedia applications: Lessons from research. *Museum Management and Curatorship, 17*(2), 173-187.

Fahy, A. (1995). *Information, the hidden resources, museum and the Internet.* Cambridge: Museum Documentation Association.

Mayer, R.E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*(2), 240-246.

Mayer, R.E. & Gallini, J.K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*(4), 715-726.

Nonaka, I. & Konno, N. (1998). The concept of Ba: Building a foundation for knowledge creation. *California Management Review, 40*(3), 40-54.

Pearce, S. (1995). Collecting as medium and message. In E. Hooper-Greenhill (Ed.), *Museum, media, message*. London: Routledge.

Piacente, M. (2002a) Multimedia: Enhancing the experience. In B. Lord & G.D. Lord (Eds.), *The manual of museum exhibitions.* Altamira Press.

Piacente, M. (2002b). The language of multimedia. In B. Lord & G.D. Lord (Eds.), *The manual of museum exhibitions.* Altamira Press.

Sayre, S. (2002). Multimedia investment strategies at the Minneapolis Institute of Art. In B. Lord & G.D. Lord (Eds.), *The manual of museum exhibitions.* Altamira Press.

Spencer, H.A.D. (2002). Advanced media in museum exhibitions. In B. Lord & G.D. Lord

(Eds.), *The manual of museum exhibitions.* Altamira Press.

Von Foester, H. (1984). Building a reality. In P. Watzlawick (Ed.), *Invented reality.* New York: WWNorton & C.

Von Glaserfeld, E. (1984). Radical constructivism: An introduction. In P. Watzlawick (Ed.) *Invented Reality.* New York: WWNorton & C.

Wallace, M. (1995). Changing media, changing message. In E. Hooper-Greenhill (Ed.), *Museum, media, message.* London: Routledge.

Watzlawick, P. (Ed.) (1984). *Invented reality.* New York: WWNorton & C.

Weick, K. (1995). *Sensemaking in organizations.* Thousand Oaks, CA: Sage Publications.

KEY TERMS

CBT: Computer based training; training material is delivered using hard support (CDRom, films, and so on) or on site.

Cognitive Theory: Learning as a sense-making activity and teaching as an attempt to foster appropriate cognitive processing in the learner.

E-Learning: A way of fostering learning activity using electronic tools based on multimedia tecnologies.

Information Delivery Theory: Teaching is just a delivery of information and students are just recipients of information.

Leader-Led E-Learning: Electronic learning that involves an instructor and where students can access real-time materials (synchronous) via videoconferencing or audio or text messaging, or they can access delayed materials (asynchronous).

LO: Learning objects; single, discrete modules of educational contents with a certain goal and target, characterized by content and a teaching method that foster a certain learning tool: intellect, senses (sight, heard, and so on), fantasy, analogy, metaphor, and so on.

MMT: Multimedia technologies; all technological tools that make us able to transmit information in a very large meaning, leveraging the learning power of human senses and transforming information into knowledge stimulating the cognitive schemes of learners.

PST: Performance support tools; software that helps students in performing a task or in self-evaluating.

Self Paced E-Learning: Students access computer based (CBT) or Web-based (WBT) training materials at their own pace and so select what they wish to learn and decide when they will learn it.

Space Constraints: All kind of obstacles that arise costs of transferring from a place to another.

Time Constraints: It refers to the need to catch something just when it takes place because time flows.

WBT: Web-based training; training material is delivered using the World Wide Web.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 737-741, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.6
Teaching Computer Graphics and Multimedia:
A Practical Overview

John DiMarco
Long Island University, USA

ABSTRACT

This chapter defines and examines situations, problems, and processes faced by teachers of technology. Based on experience teaching at three colleges with different economic, academic, ethnic, and financial attributes, this chapter provides stable and practical approaches to solving common issues. Each school environment and student population presents different technical and interpersonal challenges. Real experiences involving set up of college laboratories, development of digital curriculum, and creation of accredited programs are highlighted and transferred into tangible strategies. If you are new to teaching digital subjects, this text may help you get started. If you are an experienced teacher, this may bring you a new strategy or perspective. Ultimately, this chapter aims to assist student teachers, experienced teachers, artists, information technologists, and computer scientists in becoming stronger in transferring knowledge and skills in the digital realm. In addition, the chapter hopes to invite scholars and educators to explore teaching computer graphics and multimedia within the context of their own disciplines.

INTRODUCTION

The teaching of technological output requires every student to get value from his or her experiences within the laboratory environment. The coursework and laboratory work that challenges a student should simulate real conditions. Course problems should present both conceptual and technical challenges to students. As a digital design professor, I feel that all "digital teachers" have a great responsibility to students. We must transfer knowledge and skills at the highest levels. We must be thorough in our approaches and precise in our criticisms. We teach what we know and must know what we teach. We must teach using real-world materials and techniques. Although we may have varied control over curriculum directions, we do have control over our

success or over that of our students. We must encourage participation, communication, responsiveness, and critical thinking about design and final output. We must always encourage and never insult. We must facilitate practice — and plenty of it. We must be lifetime learners. We must have a personal technology budget. We must be empathetic toward the problems of the individuals we teach. We must take responsibility for the success of the students in our courses. We must care and commit to excellence.

MOTIVATION AND CONFIDENCE

Clear Your Mind and Prepare for Intense Mental Challenges Ahead

Go into the teaching environment with a clear head and focus on the task at hand. It is nearly impossible to communicate effectively when you have worries or problems on your mind. You can use concentrated breathing right before class to free your mind and body of negative energy. Here's how:

1. Sit down (preferably on the floor or on a comfortable chair)
2. Raise your arms and breath deeply with big inhales and exhales for seven to 10 repetitions
3. Concentrate on only your breathing, repeat raising your arms, and take five deep breaths
4. Relax your mind and empty it of all thoughts
5. Get up slowly and focus on the task at hand
6. Review and talk out your lesson outline before leaving the house
7. Arrive at class a few minutes early — rushing will raise stress levels

Remember that teaching computer classes is a mental workout. It makes your mind move rapidly through a barrage of material. It requires you to take information and present it in a hierarchical structure. That requires clear thinking! Here are some tips to help build confidence and success.

Walk Around the Class Frequently

If possible, make it a point to stand briefly near each student. Doing this generates attentiveness and immediately revives a bored student. Mobility keeps people awake!

It is essential to engage students freely and often. It induces interactive communication and results in illustrating your teaching style as a "hands-on" approach. Stay in tune with student work to nurture revision and improvement. Take pride in the project's success: the student is a product of your guidance and cultivation.

How do you stay mobile? Wear comfortable shoes — you will be on your feet. And make a 10-minute rule. If you are seated for 10 minutes or more, get up and walk around. It will keep you sharp and the class listening and alert.

Avoid Frustration

Frustration is one of the biggest obstacles for computer teachers. You must be able to reverse negative student attitudes. You must be able to deal with people who, at first, do not get it. No matter how much you repeat, they do not understand and cannot execute. Try to refrain from berating foolish questions. Be critical, but kind. Become the nicest person you can be — then get even nicer. Being friendly results in relaxation and comfort for the student. Be friendly, but firm. Expect results and anticipate initial failure. Be ready to console and control the negative situation. This is extremely important in the arts where creativity must flow freely. Whoever said "patience is a virtue" never taught computer-based coursework. Do not crack under pressure. Being a computer

teacher requires the patience of a saint. Why? Because every time you have a class that exceeds one person, you have students who have different levels of knowledge and skills with computers. Unless you are teaching an introductory course in computer basics with a room full of blank slates, you will have some students who are more knowledgeable than others. Students may walk into class who have never have touched a computer. In that same class, there may be students who have used computers at home, at school, at work, or in prior training and coursework. You may have a disruptive student or an extra-experienced one who wants to stump the teacher with outlandish or unrelated personal equipment questions that cause class flow to be halted. In the classroom, your brain will be torn in many directions. Do not give up. Stay with it.

Remember Who Controls the Class — You Do!

You control the class, the class does not control you. Therefore, you should set a pace appropriate for your teaching speed — not appropriate to the levels of one or two students. Do not try to cover an abundance of material to satisfy one advanced student. You will end up losing the rest of the class. Let the student explore advanced techniques on their own time or during individualized help sessions with you. Do not lower the goals of the class to cater to one student. We work in a competitive world where computer art and multimedia may not be for every student. Let students understand the responsibilities involved in learning. Both the teacher and student are responsible for success. Without mutual effort, respect, and extensive communication learning will be retarded. Allow a lagging student to work under less intensive criteria; in an academic setting, however, grade accordingly. Sometimes it is best to allow an advanced student to explore new concepts and techniques. Encourage the right students to accelerate or to slow down while keeping the entire class working on the same project at the same pace. This seems like an impossible equilibrium, but try to achieve it. This helps when demonstrating techniques. It helps eliminate premature and old questions.

STUDENT INTERACTION

Motivate Students by Relating Hard Work to Tangible Reward, Not to Academic Failure

Be a coach. Install a winning attitude in every student! Make your class a win-win situation for all students. Let them know that the efforts they put in during class sessions will reap rewards later in their careers. Equate effort to success. Learning computer graphics and multimedia requires a concerted effort toward success. This must include diligent work inside and outside of the classroom and the computer lab. It requires a commitment to lifelong learning and the expansion of skills that occur regularly throughout one's career. Students have to buy into the fact that you are not just their teacher, but also an expert and a valuable resource for them. They have to trust your abilities and they have to share a common goal with you — their success. I believe that a teacher on the college level should try to avoid threatening students with bad grades; rather, make students aware that their grades rely upon success in the course and the projects assigned. Students' success is directly related to effort inside and outside the classroom. Here is my philosophy — I share these thoughts with my students on the first day of each new course:

To learn our craft, we must practice. Practice takes the form of projects. Projects must be completed to succeed in the courses. Completed, well-done projects in our portfolios and backgrounds help us get new projects and jobs. Projects portray our skills, abilities, and experiences. Make your

portfolio scream of completed projects and you will succeed in getting the opportunities you want.

Determine "Can't Do Students" from "Won't Do Students"

Throughout my career, different types of students have challenged me in valuable ways. I have had to learn how to guide them to meeting goals. I categorize notably challenging students in two ways: "can't do students" and "won't do students".

The reasons behind both types of students may be abundant. However, I have found that there are general consistencies between each type. Determining student type and motivation will help you in your goals to help these students succeed in the course, and inevitably to build confidence, skills, knowledge, and work.

"Can't Do Students"

- May have personal problems that are affecting class involvement (missing class for work or illness)
- May not have the time to devote to practice and/or to continued learning
- May not have the resources (lacking a computer at home or the proper software)
- May lack prerequisite skills
- May have a learning disability

"Won't Do Students"

- Put obstacles first and success last
- Always have a list of excuses for not practicing or completing projects
- Battle tooth and nail before joining the pace of the class
- Ask meaningless and obscure questions to help interrupt the flow of the class
- Venture off on their own, not following class procedures and lectures
- May feel they have superior knowledge or experience over the instructor and feel the class is a waste of time for them

Sound familiar? If you have taught computer graphics, you may have encountered these challenging students. We will explore how to make them successful.

Talk to Students and Give Out Initial Assignments to Discover Who is a "Can't Do" or a "Won't Do" Student

A little conversation and some simple questions will go a long way when trying to determine "can't do" students from "won't do" students". I want to learn about my students so I can identify potential highlights and possible problems.

Initially, get to know your student's approaches to his or her own career. What do they plan to do when they graduate? What medium and technology are they most interested in? Ask about prior classes, teachers, outside experience, and determine their genuine interest in the subject. If enthusiasm is low or if the student begins to complain about a former teacher, do not embark on the past obstacles — instead, help become a catalyst for change. Reassure the student that you are committed to their success in this course. Explain that with cooperation, trust, and collaboration, everyone wins. Let the student know that you have the confidence that they will succeed. Always avoid negative comments or confrontation.

An effective method to begin to identify certain student types is to assign a simple homework problem that involves brainstorming on paper. This can be applied to projects including database development, programming, graphic design, multimedia, web design, and video production. I usually make each student begin to write out concepts, metaphors, interaction, messages, and prospective images for the term projects. I require everyone to do it. Those students who do not complete the assignment can be assumed to be possible "won't do" students and will require extra encouragement and more one-on-one time during the learning process. You must continue to work with "won't do" students to help facilitate

involvement. Sometimes "won't do" students will require extra attention and extended feedback on projects to continue valid participation.

When you encounter "won't do" students who insist the projects are below their expectations or abilities, allow them to develop their own approaches. However, specify that they are subject to your approval. This will allow you to monitor the situation to insure the proper focus is maintained. Nevertheless, it will also allow the students to explore new and advanced territory. The result could be extraordinary work.

"Can't do" students need extreme focus to succeed in the digital realm. They will be identified during the first or second project. To help them succeed, practice time must be multiplied, hands-on instruction should be increased and, in extreme cases, students may need to be asked to explore prerequisite courses, possibly more than once. Because "can't do" students often require extra time to complete projects, this will inevitably disrupt the flow of the course during the term. Try to keep all of the students on schedule. Encourage and assist slower students to catch up by increasing practice time and simplifying project approaches and, perhaps, possibly criteria.

Develop an Approach for Each Type of Student: New, Experienced, Young, and Older

Embrace the new challenge, understand it, and conquer it. Your student populations will be a collaboration of diverse personalities, backgrounds, skill sets, dispositions, and problems. We have to embrace this fact and figure out how to cater to everyone's uniqueness in the classroom. Sounds daunting, I know. It's one of the hardest challenges to teaching digital curriculum — managing people and personalities to encourage creativity and generate results in the form of projects.

Controlling Unfavorable Situations

You must control the situation, but never become confrontational, regardless of the student's tone . Always be serious, but never angry. Listen first to understand, then explain clearly to be understood. If you feel that you are out of control, take a class break and calm down. Then discuss the incident with the person after class so that other students are not involved. This helps avoid feelings of embarrassment for you and for the student.

Your goal is to make all parties involved happy. Convey the fact that you always want a win-win situation and the student to be ultimately happy and satisfied. Show the student that you are there to help, not hinder. Let the student know that you are a resource for them to utilize. This type of giving approach will help generate trust, admiration, and connection between you and your students. Above all, be empathetic to the student's problems and concerns.

Make Proper Ergonomics a Priority for You and Your Students

Everyone will be happier and healthier when they sit properly. Ask students to sit, put their feet flat and have both hands ready to use the computer. Institute a "two hands at all time rule" because it will keep students alert and working properly. You cannot slouch at a computer and use two hands. It is impossible. Should you request that students use two hands on a computer? I explain to students that in the "real world", (I refer to the "real world" often), creatives are paid by the time they spend on projects. All programs have shortcut keys (such as "apple + z" or "control + z" to undo) and all graphics and layout programs have multi-key combinations that are needed for performing certain functions (for example, to scale an object and keep it's aspect ratio, you must hold shift in most programs). Therefore, using two hands becomes the professional approach that everyone should emulate.

Ergonomics is the study of the human body at work. As a science, it has roots in the industrial revolution. Initially, time and motion studies were used to improve workers' performance. Today, computing and medical societies realize how important it is to health.

Besides efficiency, maintaining sound ergonomics helps to eliminate physical stresses that working on computers can foster. Pain in the wrist, forearm, back, neck, and legs can be lessened or eliminated by employing sound ergonomic components including a drawing tablet, a good chair, proper posture, adequate lighting, and an appropriately-sized workspace. These assets may not be present in your teaching environment. You should try to obtain as many as possible because they make a difference in the quality of work and student success.

TEACHING STRATEGIES FOR BUILDING SUCCESS

Understanding Your Objectives

The first question that must be clarified before you start to teach is, "What are my objectives?" What do I want to accomplish? What goals are to be met? There are certain state standards that must be reached when teaching in a public school setting. In college there are certain standards for each department and there are certain requirements that must be enforced for institutional accreditation. With all of those standards and objectives, why do we need anymore? The objectives discussed here are important to the success of a teacher of computer graphics and multimedia in any discipline. I stress the word success. In my opinion, this list has no particular order.

- *Objective:* Introduce new techniques and technologies to the student.
- *Objective:* Develop style by discussing historical and professional references.
- *Objective:* Adhere to your prescribed "state standards" if they dictate curriculum.
- *Objective:* Teach the student new vocabulary related to the technology and design process.
- *Objective:* Develop student understanding of the design process and its relationship to solving problems and completing projects in their discipline.
- *Objective:* Develop "computer confidence" in the student.
- *Objective:* Motivate the student to cultivate new creative concepts, approaches, and content into their work.
- *Objective:* Transfer practical skills and pass on real world experiences to students.
- *Objective:* Direct the student towards completion of a tangible project that is portfolio-ready.
- *Objective:* Critique student work and offer critical feedback and guidance on design, usability, visual, presentation, and commercial strength.
- *Objective:* Present student work to a broad audience (web portfolios).

Focusing on these objectives will help you and your students succeed. Remember: begin with the end in sight. You want your students to finish the course with a new interest in the subject, exposure to jargon and vocabulary words, confidence that they can use computers effectively, and an understanding of the design process within their discipline. You also want the student to develop a new skill set, portfolio- ready projects, and an understanding of how feedback can improve project strength.

Teach in Outlines

Teaching in outlines presents information in a hierarchical structure. Structure allows students to grasp the information in steps. Each step is a building block to the next step. Without the

hierarchical structure, there is confusion on appropriate direction and process.

For example, students in a digital imaging class are using Adobe PhotoShop. What tools or techniques are taught first? What project will be issued to reinforce the use of those tools and techniques? Should digital painting techniques or image manipulation be taught first? What about production using layers and paths? What do we lecture about on the first day? Development of a hierarchical structure is vital to teaching computers, graphic design, and multimedia.

Here's how to teach in outlines. First, perform research. Acquire books, articles, and web content on the technology and curriculum samples or cases. Go to publishers' and booksellers' websites to get abstracts and table of content samples. Read them to discover and learn about the things that are unfamiliar to you. Create a simple list of five major areas that seem important to the technology. Create an outline of these five topics. Write down important information regarding the technology and the applications. Specifically, understand what tools and applications are used to execute projects to completion. Remember to look for hierarchical structure. See how the various books and articles describe use of the technology. Take note of chapters and content. Does the book seem to move from beginner techniques to advanced? Examine other digital curriculums related to your discipline to expand your thoughts on subject hierarchy. Once you feel comfortable understanding the flow of information and the design process needed to complete a project, document it in a lesson plan. Then follow the steps in the plan to completing the process and complete the project. Alternatively, jump around through the table of contents of books to scrutinize the proper hierarchy to fit your lesson plans. Keep it simple. Don't overload your students or yourself with too many concepts too fast. Stick with the hierarchy and the outline. Your major topic outline will help you to figure out what to teach without overwhelming you with too much information. This is crucial to learning in small steps. Small steps in learning will ultimately develop a complete understanding of the subject.

Use a Topic Outline

The topic outline should include this information:

1. *Provide an overview of the technology to the class.* Explain what medium the technology is used for (TV, computer kiosk, video, gaming, print, etc.). Provide the names of the companies who make the applications. Give an overview of the class projects that will be created using these tools and technologies.
2. *Present some new, vital vocabulary words.* These will come up so you need to clarify them early. Explain resolution to a digital imaging class or vector graphics to a digital illustration class on the first day.
3. *Describe the tools, palettes, and menus in overview form.* Use metaphors to relate the fine art world with the digital art world. For example, talk about Adobe PhotoShop having a canvas, a palette, and brushes all for painting — digital, of course.
4. *Describe the creative design process* that the class will go through to develop concepts, source materials, and content for the projects. Plant seeds in students' minds so that they will have creative control and consistently generate new approaches to their art, design, programming, multimedia, and communication pieces.

Talk About it First to Get Feedback and Levels of Experience

Always discuss the subject before everyone jumps to use the computers. Make your discussion interactive. Don't just lecture — question. Make a strong appeal to have the entire class

give answers. What are you asking? Questions that you have devised from your outline. These questions will help you gain an understanding of the experience levels of your students. It will also help generate some background information on the students that may come in helpful later. Shy students, talkative ones, highly experienced, inexperienced, and exceptional students will all coexist within your learning environment.

Have a general discussion that involves the class. Begin with a question directly related to the name of the course. WHAT IS DIGITAL VIDEO? What does digital mean? Does our VCR play Digital Video? These are the questions I ask eager students. Who has heard of Adobe Premiere or Apple Final Cut Pro? What do they do? Who in the class has used them?

When you ask with enthusiasm and you demand involvement, students respond. You will notice that everyone is thinking. There is no better gratification for a student in class then presenting the right answer in a class discussion. Students strive for it. They may give all the wrong answers, but that's fine at this point. The chance to give the right one is worth it. The main thing is that we have succeeded in making them think about the processes, technologies, and applications they will use shortly.

Factor in Practice Time at the Start

Good design work takes a good amount of time. Conceptualization is crucial and changes are constantly upgrading the design. Thus, it is necessary to increase our time for the project. At some point, there is a deadline. Let students know immediately that they should expect to spend time outside of class to work on projects. If the context of the class is not a serious commercial-level audience, make sure practice time is factored into class time. Students learn computer graphics and multimedia in steps. Each step leads to the completion of a technique. The techniques help build the project to completion. The techniques require practice for the ability to execute and also to remember after the class is over. Maybe the person has to go back to work on Monday and use the new skills immediately. Practice must be factored in! If students can't make a commitment to practice they will never learn to complete projects and reach class goals.

Talk About It, Show It, Do It, and Do It Again: Answer Questions and Do It One More Time

This is a process that I use constantly to achieve retention. Retention is the key to student development. If students can't remember the technique, how can they implement it? If you look at students and they give the impression that you are going too slowly, you've hit the mark: retention. There is no such thing as over-teaching a topic. Be sure everyone understands before you move on to the next topic.

Review should be part of everyday lectures. Techniques lead into other techniques. Therefore you can always lead into a discussion on a new topic by reviewing the last topic and it's relationship to the next one (remember the hierarchy of information).

Don't Overload Students with Useless Information

We should not be teaching every single palette, menu, and piece of a software package or technology. Why? Information overload and time. The student needs to learn the material following the hierarchical structure that produces project results. There are things that will be left out simply because they do not rate high in the hierarchy. The information has less value. Therefore, it does not need to be covered formally. Books and student discovery can help teach the less important features and save brain space for the most topical information and techniques.

Projects are Mandatory and Needed — Assign Them, Critique Them, and Improve Them

The most important indication that learning and growth have occurred in art, design, film, video, computer art, graphic design, multimedia, web design, illustration, motion graphics, 2d design, programming, and 3d design are the projects. Practice should come in the form of project work.

Student projects should be evaluated on an ongoing basis. The critique of the student work should provide a forum for comments that provide constructive criticism. The goal is to get comments, not just compliments. The student must also realize that what is said is opinion only and can be discounted as easily as it can be used for improvement in the art-making process. However, the design process has specifications that need to be followed to complete a project accurately. The designer cannot alter these "specs". Elements such as page size, color usage, typography treatments, media, and message are all dictated in the design process and instructions must be followed. After critique, adequate time should be given for students to refine their work and then submit it for a grade.

Critique for design projects in any discipline may include criticism on:

- *Concept:* original idea that facilitates and supports a design solution
- *Development:* research and design aimed at defining project goals and strategies
- *Technique:* skills and processes used to execute design production
- *Style:* specific approach to creating unique identity and definitive cohesion within the project design
- *Usability:* functionality and effectiveness within the media which is related to stated goals or specs
- *Output:* adoption into final deliverable (print, web, CD-ROM, Digital Video, DVD, executable application)
- *Presentation:* visual and verbal summary of the value and success of the project

Stifle Disruptive Students Early

Teaching is difficult enough without having a disruptive student. Usually disruptive students come in pairs of two. The word RUDE is nonexistent in their vocabularies. However, you must be polite and provide verbal and nonverbal signals that you want quiet when speaking and during in class work periods.

Typically, talking between two students is common. Simply and politely, ask everyone to listen to you now. That will stop talking sometimes. If it is persistent, stand near the talkers and ask them directly to wait until break to talk. Explain to them that the class is going to start and that everyone needs to work in a quiet environment. If it gets bad, speak to the student after class and explain that there are others who cannot learn with the constant disruptions. Explain that everyone is paying for this class and that talking is slowing the pace and the success of the class. This should help curb or solve the problem. The last resort: speak to your supervisor about possible solutions.

Always remember to require headphones for students working on multimedia projects containing audio.

Don't Do the Work for the Student

You have heard the proverb many times. Give a man a fish and feed him today. Teach him how to fish and feed him forever. It's the same with teaching digital coursework. The student must perform the techniques over and over again to learn. If we try to help students by sitting at their computers and doing the work, we are actually hurting them . Students will expect the teacher to fix problems by clicking a few buttons. That's not the case. The teacher should rarely touch students' computers. The exception is in extreme situations or when students have gone way beyond

where they should be and now need to have their screens reset.

I tell students that it's easier for me to fix a problem than for them try to figure out the solution based on what they know. Sometimes, I encourage students to think about what the problem is and try to find a solution on their own — then we will discuss how the outcome was derived. Thinking through a problem is part of the learning process and design process. It is also important to help students become problem solvers.

Save Early, Save Often

Institute a blatant policy on saving. Too many times I have heard horror stories and consoled distraught students after they have had digital tragedies. I've seen it all. Zip disks going into CLICK DEATH, where the Zip drive is clicking your disk and corrupting your data (Iomega was sued for this), a meltdown of media due to being left on the car seat in hot sunlight, lost files and disks, and crashed hard drives. It happens. We should be prepared with backups. The student should be saving constantly. Computers crash frequently, regardless of platform, brand, or operating system. If we have created something that is important, we should be saving it in two places. By the second or third class meeting, you should have a formal lesson on saving. You want students to understand how to save, why it is so important, and where it is done.

All students should have external storage media with them by the second class. A zip or writable CD-ROM if applicable to your workstations. Also, students can use plug-and-play Firewire hard drives that provide hundreds of gigabytes of storage space. Show the student how to access external media. Teach them how to move and delete files. Show them where they can save on the local hard drive or server if available. Make saving a priority. There is nothing worse than working on something for hours and then losing all the work because you forgot to save. As teachers, we don't want our students to experience that dilemma. It is our job to remind them frequently and to train them to save early and often.

Utilize Real World Examples

Every time I explain a topic, an example from the real world is shown. Many times it's work that I had a role in, many times it isn't. By using real world examples we justify the techniques and approaches being taught. We show students that this is something they may be doing in their careers. We also alert them to the fact that the skills and techniques they are learning are important to their technical and professional development. For example, I'll show a magazine spread and explain text wrap, alignment, typography, and design. Then I'll mention that those design skills and techniques are very important if you want to work in publishing production and design. Sometimes that will make people sit up in their chairs and listen a bit more seriously.

Provide Early Timelines for Exams and Projects

Timelines are crucial to project management. Providing a timeline to your students before the course begins is essential to getting projects and assignments completed on time. The timeline also provides a visual workload for students to factor into their schedules and lives. On the timeline, list the dates projects are due, when exams will be given, and tentatively when topics will be covered. This timeline should be part of the syllabus. The timeline will also help you teach topics efficiently and help you to not forget topics.

Always Use the Correct Vocabulary

I hate when computer teachers confuse memory (ram) with disk space. Someone says, "My zip is full, I'm out of memory." That person is not out of memory; they are lacking disk space on their

storage media. We must use the correct vocabulary. The jargon that exists in computer graphics, multimedia, the Internet, and within the world of computers is endless. We can't know it all, but we as teachers should know our discipline's variety. Understanding jargon is crucial to participating in everyday operations. If you forget the difference between Fahrenheit and Celsius and you are a meteorologist, you are in trouble. The same holds true if you are a digital production artist and you forget the difference between an .EPS file and a JPEG file. Use the correct vocabulary in lectures, notes, projects, and in answering questions. It will also raise your level of credibility as the instructor and perceived expert.

Administer Vocabulary Exams

Students need a strong digital vocabulary to interact professionally with colleagues, vendors, potential employers, clients, and industry. Understanding vocabulary allows the student to read and understand books and articles on the subject more fully. In addition, this jargon is found in the project process. Using vocabulary correctly and giving vocabulary exams are the most successful methods for grasping vocabulary.

Involve Everyone During Questions and Feedback

You are a resource for the student. Provide feedback and direction, but don't do the work for the students. Whether it is creative conceptualizing or digital production, students must explore the process. Answer questions with questions. Make certain that students understand the question and have genuinely not found the answer. Many times students will take a passive approach and ask a simple question that they should know or will ask simply to get the instructor's attention. Try to encourage learning during questions. Get the group involved in questions. Have classmates help out during question and answers sessions. It helps sharpen vocabulary words and concepts for students. When feedback is being given during critiques, be sure everyone is involved and giving their constructive comments. This process helps evoke a synergistic team approach to analyzing projects.

Build Confidence from Start to Finish

Lack of confidence is a terrible state of mind that occupies many students. Typically, older and returning students show it the most. As teachers, it's our job to instill confidence in students. How we do it requires some extra effort on our parts. Here are some things to remember when you want to build confidence:

- Don't patronize students. Never talk down to them.
- Listen first and explain second. Understand students, then teach them.
- Never allow negative talk to consume the student. Regardless of the situation, let the student know that if they put the effort in, rewards will follow.
- Be a coach, project manager, mentor, and motivator.
- Let students know that you are there to help them succeed.

Act as an Invisible Project Manager

The project manager is the person who makes sure that the project is done on time and correctly. We must act as project managers when teaching computer graphics and multimedia classes. The students we teach are going to be working on projects. There will be obstacles that hurt the project process. As teachers, we must help the student remove the obstacles that hinder project completion. Ask students what is happening with their projects. What stage are they at? What is left to do so that projects can be completed? Ask what

problems are being encountered that are stopping the progress of the project. Offer solutions and provide updates on when projects are due and how much time is left for completion.

SETTING UP YOUR LABORATORY AND CLASSROOM ENVIRONMENT

Take an Active Role in the Development, Maintenance, or Upgrade of your Lab

The way to learn about technology purchasing is to get your hands dirty in it. Dig in to catalogs and websites. Find out the names of manufacturers, vendors, and get a sense of what prices exist (the catalogs will tell you that). Developing a lab requires you to take the time to learn about the environment. Lab maintenance is a good way to keep the lab running well. You should acquaint yourself with standard maintenance procedures for each platform (especially the one you use). They can be found in your systems owner's manual. You can also use software tools such as Norton Utilities to help you maintain and repair your workstations. Learn about future upgrade items by researching. Find out early when there are new versions and tools that are important to progress. Understand the technologies early and then, when they are instituted, you will be a step ahead.

Determine Your Needs on Paper First, then Write the Purchase Order

Setting up a computer lab requires financial, organizational, technological, and logistical aptitudes. Also, it demands some insight regarding the politics and red tape present in your organization.

The best way to organize any project is to write it out. Get down to the practical aspects and develop an outline for the new or upgraded lab. Address these areas in your brainstorming session:

- What is your anticipated budget? Sometimes this number changes, so prioritize.
- What is the student maximum per course? How many courses per day?
- Which courses will be taught in the lab? Information Technology, Computer Science, Multimedia, Graphic Design, Video, or perhaps all of them?
- Will the lab need internet access? It most definitely does, regardless of the class.

Nevertheless, remember your priorities. The Internet is a necessity in some courses. Other courses may not need internet capabilities. However, all classes can benefit from using the internet for research, exhibition, or communication.

- Will the lab have to accommodate students with disabilities or special needs? This is a critical concern in some institutions. Typically equipment and furniture considerations differ in this situation.
- What are the electrical and lighting situations in the room? Will outlets or conduits have to be installed? Will lights have to be installed to improve the functionality of the space?
- Based on the courses taught in the classroom, what programs need to be purchased?
- What platform (Macintosh or Windows) does the staff prefer? What platform is dominant in the industry? What platform currently exists (if any)?

A clear view of priorities will emerge after answering the above questions. Start to prioritize the truly important items. Add the luxury items to a wish list and hope for future budget money. An example of priority items include: CD ROM burners, color laser printers, digital cameras, constantly updated versions of software, presentation panels, and internet access. Without these items, you cannot run a lab and teach an entire program.

What to Use for Graphics and Multimedia? The Age-Old Question: Must We Use Macintosh or Can We Use Windows?

I am asked this question at least 100 times per year. My answer stays the same throughout OS changes, new versions, and bad and good news. The answer is that it is up to the staff. Whatever the staff is most comfortable with is best to have. Macintosh is widely used throughout advertising, digital production, and new media. Windows 98/2000/XP is dominant in business applications, information sciences, and computer science. Running graphical software and producing multimedia on the Windows platform is now as seamless as on the Macintosh. The real question is, again, what does the staff feel most comfortable with? Moreover, what is the industry standard? There are some distinctions that bring the old folklore that Mac and Windows PC's were mainstream competitors and they did not work well together exchanging files. That has changed dramatically since about Mac OS 6 (1993). That's when Apple bundled the application PC exchange with the OS. The extension allows Windows media to be read on Macintosh systems — right out of the box. Windows does not have that ability yet. Therefore, Macintosh has been touted as the friendlier platform in some circles. Throughout the publishing industry, Macintosh has been a standard platform since the inception of desktop systems.

Three-dimensional modeling, motion graphics, and multimedia applications gave way to the need for workstations that employed multiple processors and huge storage drives. Workstations are used in television production, film effects, video, DVD and CDR authoring. The workstation allows a massive amount of processing power to output gigantic files. Macintosh systems cannot provide the muscle that some workstations can. However, Macintosh is competing in the desktop arena by providing innovative hardware and software solutions for DV and DVD production. Apple computers allow users to capture, develop, and edit full-length digital video using a Digital Video camera. The digitized video can then be output to multiple media including web, broadcast, video, CD-ROM, and DVD.

Here are the main differences between Mac and Windows from a user's point of view:

- Windows systems on a base price level can be purchased at a lower cost than Macintosh. The reason is that there are so many manufacturers of windows-based computers in the United States that prices fluctuate and there are constant price wars in the PC market. Macintosh is a brand that owns its platform. No other company manufactures products running the MAC OS. Apple products are sold at a fair market price and are very competitive with comparably matched Windows Systems. However, a few operational differences make Macintosh more desirable for the print, multimedia, and content creation arenas.
- Macintosh accepts and reads files on the Mac platform and will open Windows files created with the same application. Windows will do the same, but will only read Windows media.
- Macintosh allows digital video input and output right out of the box. Previously, many Windows computers could not say that. If you want Firewire (ieee1394 or Ilink) technology for digital video use on the Windows platform, you must buy an aftermarket video board. You must install it and hope it is compatible with your computer and digital video editing software. Most PC manufacturers are adding Firewire cards bundled with their systems to avoid incompatibility issues. Macs also typically come with digital video editing software when purchased. It's not the high desktop applications we use in the professional industry, but is enough to create, edit, and output digital video — right

out of the box. It's great for educational settings where content and basic knowledge are more important than teaching high-end applications.

- Macintosh computers have network capability directly out of the box. All that is needed is an Ethernet hub and Cat 5 cables and the AppleTalk takes care of the rest. For Windows machines, an Ethernet card would have to be purchased for each computer on the network as well as a hub. The Macs come with Ethernet built right in.

Bottom line, it does not really matter. Windows computers may lack some simple features that we have come to adore on Macintosh, but when it comes to sheer horsepower, Windows XP multiprocessor workstations are the choice for the professional film and TV editing, three- and four- dimensional modeling, and animation developers. The price factor makes Windows a bit more desirable, but Macs give added features out of the box that provide networking and multimedia capabilities to students. In a digital design lab environment, Macintosh may provide some functional advantages. In a computer science environment, Windows machines are typically more desirable. Ask your colleagues and staff what they prefer. Then ask them why. Collect the information and make a majority decision based on budget and priority.

Prioritize Your Purchase Regardless of the Platform and Put First Things First!

You must have your priorities in order when you are developing a computer lab. The decisions you make when the lab is delivered are the same decisions that come back to haunt you when you need more resources or things do not meet expectations.

- *First priority:* Make sure that you have enough computers for every student. If the class size exceeds the number of computers due to enrollment, cut class size. If class size is too large, you will need to make shifts of student to work on stations. Your teaching load will double per class because you will be shuffling around students to get everyone working on something. Inevitably, student work will suffer. There is a better way. Demand there are enough computers for each student.
- *Second priority:* Get as much ram as you can afford. Load up. You'll be happy when software versions change and your hardware budget is on hold until further notice.
- *Third priority:* Removable storage drives such as Iomega ZIP drives. Students need to back-up and transport their work. These drives allow them to do that. Having one drive becomes chaos. Saving and archiving become afterthoughts and hassles to the students. Also a good choice, but a bit more expensive for the student compared to a ZIP disk are removable Firewire hard drives. Eliminate the fear of losing files by making student backup an important priority. Also needed are archival media and drives, including CDR and DVD. These drives should not be used for primary, daily storage, but for final project output and archiving.

If You Develop Any Computer Lab Learning Environment, Try to Make the Following Items Part of Your Proposal

These will make teaching much more manageable. I'm sorry if these items are considered luxuries due to budget constraints. However, I cannot stress the importance of these items on what I call "quality of lab teaching life".

- *Hardware security system* including cables and padlocks for systems, monitors, and peripherals to keep the lab safe from theft.
- *Software security system* to lockout vital folders like the System Folder. These are now the duty of the lab manager. The needed functions can be found in the latest network operating systems for each platform.
- *A presentation panel for display on a screen or wall.* Although expensive, these technology learning tools are extremely helpful in the digital (smart) classroom. A decent one will cost you $3,000 to $4,000 dollars. Make sure you consider this item seriously when distributing your budget.
- *Pneumatic adjustable chairs.* This should be nonnegotiable. Bad chairs breed bad work habits. To help ensure an ergonomic lab environment, you should insist on the best chairs you can afford.
- *A file server and an Ethernet network.* This will help you transfer files and applications between stations and will allow maintenance to be easier with all machines connected. You will need this for internet access throughout your lab and to institute a software security system. Add a tape backup to the configuration for complete backup of the server.
- *Removable, rewritable, cross platform storage device* such as a Zip drive. Students need these to backup and transport files.
- *Internet accesses via a high-speed line or backbone,* especially if you are teaching a web centered course. A T1, T3, or cable connection is the minimum.
- *Server space and FTP access for web classes.* You should demand this if you are teaching a web design course.
- *A scanner for image acquisition.* One is the bare minimum you will need. This is a necessity. Without it, content will be virtually nonexistent in digital imaging and layout courses. Also include a digital camera for shooting stills and small video clips and a digital video camera for capturing full-length digital video and audio.
- *Enough computers for all students to have their own workstations during class time.* Without this, you are really challenged. It is not an impossible situation, but it requires some compromise to your teaching schedule. Inevitably, students will suffer. You can't watch and learn computer graphics and multimedia. You have to be hands in and knee deep, practicing constantly.
- *Lab hours outside of class.* This will allow students to practice and work on projects outside class time. Even if they are limited to small increments, lab hours are necessary for student abilities and confidence to grow. There has to be someone in charge during lab hours, so think about work-study students or graduate assistants to help with lab management.

CONCLUSION

Building student confidence, developing project- based skills, presenting vocabulary, and working towards project -based goals are crucial components in helping students succeed in digital coursework. But before you can do it, you must become comfortable with not knowing everything and understanding that you will be growing perpetually. Becoming an expert at teaching digital subjects presents very demanding challenges. Understanding, adapting to, and conquering those challenges will be realized through perpetual research and raw experience.

The future will undoubtedly present more and more interdisciplinary scenarios for programmers, designers, artists, production professional, writers, and musicians. This surge towards collaboration will be reflected in digital education. In the past decade, multitudes of schools have instituted new programs in multimedia, interactive multimedia, new media, educational technology, informa-

tion technology, instructional technology, and many more multi-discipline disciplines. The convergence of media, process, skills, and deliverables makes teaching computer graphics and multimedia an extremely challenging, dynamic responsibility that requires artists to learn more programming and programmers to learn more about visual communication.

REFERENCES

Heller, S. (1998). *The Education of a Graphic Designer.* New York: Allworth Press.

Heller, S. (2001). *The Education of an E-designer.* New York: Allworth Press.

Michalak, D. F., & Yager, E.G. (1979). *Making the Training Process Work* (7-72). New York: Harper & Row Publishers.

Tieger, P. & Barron-Tieger, B. (1988). *The Art of Speedreading People*. Boston, MA: Little, Brown & Company.

This work was previously published in Computer Graphics and Multimedia: Applications, Problems and Solutions, edited by J. DiMarco, pp. 1-23, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.7
Evaluating Learning Management Systems:
Leveraging Learned Experiences from Interactive Multimedia

Katia Passerini
New Jersey Institute of Technology, USA

ABSTRACT

This paper maintains that the use of multimedia content in Web-based instruction — facilitated by the proliferation and standardization of learning management systems (LMS) — calls for the extension of traditional multimedia design and evaluation guidelines to the Web. The compliance with these guidelines needs to be thoroughly evaluated by any institution using (or planning to use) Web-based learning management systems. In addition to providing criteria and examples for the evaluation of these systems, the paper includes a survey instrument that can be used for university-wide assessments of the design effectiveness of technologies that support learning. As an example, the proposed evaluation instrument is applied to a learning management system developed at a large university in the United States. While the assessment refers to one system, the model, the instructional and design evaluation criteria, and the questionnaire are built for use in any organization conducting a formative and summative evaluation or a selection of learning technologies.

INTRODUCTION: LEARNING MANAGEMENT SYSTEMS

Learning Management Systems (LMS) are Web-based applications that support online teaching or supplement face-to-face instruction. Typical functionalities of LMS include Web course design, Web course collaboration tools, and Web course management (Hall & Hall, 2004; Hills, 2003c).

The **course design** features provide templates for course organization. Instructors control the content and have some impact on the screen layout (changing features such as color and screen placement). Students can post information on personal Web pages or can create areas to post

assignments and discussion topics. Search tools are available for quick access to materials.

The **collaboration tools** include synchronous (chat) and asynchronous components (discussions areas similar to listservs). Faculty can use bulletin boards to post course-related announcements. Electronic messaging within the LMS provides a repository for course-related messages. Whiteboards are used especially with mathematical and visual information. File sharing and workgroups are particularly useful for team-based activities enabling simultaneous file editing by several users.

The **course management** features enable student grading, performance tracking throughout the course, and the calculation of time spent using the software applications. They also enable instructors to design online quizzes, randomize questions from a database, and assess response time.

In addition to the above, a number of administrative features provide security and technical support for faculty and students. Table 1 lists typical LMS areas contained in many commercial and open-source applications such as WebCT, Blackboard, and Lotus LMS.

EVALUATING LEARNING MANAGEMENT SYSTEMS

Stoner (1996) defines a learning technology as any application of technology for the enhancement of teaching, learning and assessment. This definition includes the use of network communication systems and embraces a large number of multimedia and Web applications. Learning management systems that enable classroom instruction on the Web and/or support face-to-face instruction with access to online learning repositories of course materials fall within this definition of "learning technology." When integrating a learning technology into a traditional curriculum, a thorough evaluation of its key design and instructional characteristics is a critical element and a pre-requisite for its successful implementation (Bersin, 2005; Hills, 2003b).

Stoner proposes a system design approach to the integration of learning technologies into traditional courses (or sections thereof). This approach draws on methodologies widely used in the design and implementation of computerized-information systems (Lucas, 1994; O'Brien, 2005) and in systems approaches to instructional design (Gagné, Briggs, & Wagner, 1988). Stoner's model suggests a careful data collection on course type, students, and resources available. He encourages

Table 1. LMS Features

Course Design Features	**Collaboration Tools**
Instructor-centered sample course	Discussion options
Course templates	Asynchronous/threaded
Search tools	Synchronous (chat)
Student home pages	Chat sessions logs
	Bulletin board
Course Mgt. Features	E-mail
Student grading	File sharing
Student tracking	Whiteboard
Assessment tools	Workgroups
Timed quizzes	
	Administrative Features
	Security
	Tech support

the research of alternative solutions: "These will need to be formulated in some detail, identifying the learning technology [courseware] to be used and how it might be used and integrated within the course(s) being considered."

This paper presents a framework for the assessment of LMS leveraging the system design approach suggested by Stoner. Particularly, it relies on lessons learned in the design of interactive multimedia. It applies the design evaluation criteria on a specific LMS developed at the George Washington University, the Prometheus system, to introduce a specific example of the evaluation protocol here presented.

Types of Evaluations

There are several approaches to conducting evaluations (Johnson & Ruppert, 2002; Hills, 2003b). Ideally, several types of evaluations should be implemented. In reality, financial, temporal, and human resource constraints limit the options (often in favor of "late" summative evaluations). Four main approaches to evaluating learning technologies are seen in Table 2.

Formative Evaluation

Formative evaluation is testing conducted on selected samples of the user population while the product is still being developed (prototypes). Formative evaluation use open-ended methods, survey questionnaires, or confidence logs (users' self-assessment of their knowledge). The key constraint of this method of evaluation is its timing. Authors describing the planning efforts of formative evaluations note that it is difficult to plan and implement testing early enough so that changes can be made (Alessi & Trollip, 1991). Often, resource constraints do not enable the administration of formative evaluations.

Summative Evaluation

Summative evaluation is a process that concerns the final evaluation. This evaluation usually focuses on the user (rather than the application) because it is conducted after the product release. It is used to inform decision on future developments, as a product review, and as a user-satisfaction data collection instrument. Traditional surveys or assessment tests can be used, as well as observations, interviews, and other qualitative and quantitative data.

Illuminative Evaluation

The aim of illuminative evaluations is to discover what factors and issues are important to the participants in a particular learning situation, which may differ with the developer's judgment. Draper, Henderson, Brown, and McAteer (1996) state that "illuminative evaluation has a systematic focus on discovering the unexpected, using approaches inspired by anthropology rather than psychology," and that these approaches have a significant effect on the users.

Integrative Evaluation

Integrative evaluation is "aimed at improving teaching and learning by better integration of the

Table 2. Types of evaluations

Evaluation Type	Purpose
Formative	To help improve the design (conducted during development)
Summative	To assess the product and its functionality (conducted after release)
Illuminative	To uncover important factors latent in applications
Integrative	To help users extract all the benefits of a learning technology

learning technology into the overall situation. It is not primarily either formative or summative of the software, as what is both measured and modified is most often not the software but surrounding materials and activities. It is not merely reporting on measurements as summative evaluation is, because it typically leads to immediate action in the form of changes" (Draper et al., 1996). For example, if all the students in a classroom complain about the use of the technology for a particular learning outcome, the instructor and the developers need to reevaluate the tool and its current application. "Is the feature able to achieve its intended purpose?" If it is not, the developers should promptly modify the system, based on user feedback.

This paper focuses only on the first two types of evaluation: the formative and summative models. The scope of the discussion is limited to two models in consideration of space limitations and to present examples of LMS evaluation models actually run at a large university. It describes the content and assessment procedures for applying these evaluations to learning management systems. The paper draws from the interactive multimedia literature to identify frameworks and criteria for LMS assessments.

CRITERIA TO EVALUATE LEARNING MANAGEMENT SYSTEMS: BORROWING FROM INTERACTIVE MULTIMEDIA

Interactive multimedia instruction is traditionally grounded in several years of experience with interface design, human-computer interaction and computer-supported mediated learning. In the mid-and late nineties, instructional multimedia was partially replaced by LMS systems (Taylor, 2003). New applications competed in emerging as tools to facilitate transfer of in-class materials to the World Wide Web. Initially, these systems failed to leverage the design lessons from interactive multimedia because of the then clear hiatus between the Web and multimedia systems. Today, pervasive broadband access has brought about the possibility of delivering multimedia content in a Web space with a relatively low bandwidth impact. Finally, several multimedia applications have started to be transferred online (Watson & Hardaker, 2005) through Macromedia Flash and Java programs, lowering the gap between interactive multimedia systems and LMS.

The distinction between interactive multimedia and the Web is becoming "blurry" (Hedberg, Brown, & Arrighi, 1997). If interactive multimedia was perceived to be bound to the shell of a physical container (the cd-rom), today's online delivery capabilities enable hyper-linking and navigation as in a Web-based system. And where interactive multimedia systems are still constrained by the boundaries of a self-contained application, also LMS suffer from the same limits. As in interactive multimedia products, LMS rely on a self-contained (online) shell within which both instructors and students (and only the instructor and registered students) coherently navigate to organize and retrieve documents (within the available templates) (Hall & Hall, 2004). In this context, the areas of coincidence among interactive multimedia design and Web-design guidelines increase. As the coincidence grows, lessons learned and validated interactive multimedia frameworks can be leveraged to evaluate the effectiveness of an LMS (before its curriculum integration) (Coates, James, & Baldwin, 2005).

As discussed next, an examination of multimedia and Web development principles furthers these statements eliciting a close mapping of interactive multimedia design guidelines with Web-based instructional systems design guidelines. It extends multimedia design principles (for example, Reeves, 1993) to LMS models.

Designing multimedia for instruction requires major attention to two main factors: coherence and cognitive load (Yi & Davis 2003). Coherence of screen design is a key element of comprehension

as it facilitates the construction of the mental models for the learner. The higher the coherence, the easier it is for the learner to comprehend.

- **Coherence** needs to be reached at a *small scale,* linking pieces of information together for local coherence, and on a *large scale*, reminding the user about the relationships between the current screen and the learning domain.
- **Cognitive load** is defined as any effort in addition to reading that affects comprehension (i.e. navigation efforts or adjustment to the user interface). The higher the cognitive load the more difficult it is for the learner to comprehend. Strategies for reducing the cognitive load include creating a good balance on "distance", "focus", and "proportion" (Szabo & Kanuka, 1999). For example, key elements on the screen can be placed or given a different layout or shape based on importance. Cognitive load is also reduced by using clear navigational strategies. For example, hyperlinks/buttons need to be user-friendly and easily understandable also by a novice user.

Multimedia applications follow screen design, navigation and interactivity design guidelines that are informed by cognitive load and coherence principles:

- *Ease of use*. The perceived ease/difficulty of user interaction with a multimedia program: the more intuitive the application user interface, the less impact on the user cognitive load.
- *Screen design*. Screen design in multimedia relates to the coordination of text and graphics to present a sequenced content. This content facilitates understanding (Mukherjee & Edmonds, 1993) with each screen providing effective instruction, appropriate navigation tools and pleasing design/visual aesthetics (Milheim & Lavix, 1992). Each screen must display a **navigation toolbox** at the bottom, **title and instruction** areas at the top of the screen, with the **body area** containing media clips in the center (Stemler, 1997).
- *Information presentation*. Visual clues and information on the screen cannot be cluttered: too many representational clues (icons) or too much declarative text in one screen creates confusion and overwhelms the user (Overbaugh, 1994).
- *Level of interactivity.* Interactivity is a key distinctive feature of interactive multimedia and should be provided frequently, at least every three or four screens (Orr, Golas, & Yao, 1994).
- *Navigation*. Navigation should occur through simple interfaces, using facilitating metaphors and familiar concepts (Gurak, 1992). The icons should clearly show whether they are hyperlinks to other screens (by color, form, or mouse-over effects).
- *Quality of media/media integration.* Individual media (text, sound, video, and animation) within a multimedia application need to be synchronized based on content, space, and time of the animation.
- *Mapping.* Thuring et al. (1995) suggest several hyperlinking guidelines. They include the clear identification of hyperlinks, the visualization of the document structure, the inclusion of navigational tools, and so forth.

Web design guidelines are closely related to design principles identified in interactive multimedia. In the following list, design principles identified by Jones, Farquhar, and Surry (1995) are mapped to interactive multimedia guidelines. For example, well-designed **Web-based application systems** need to:

1. **Provide structural clues:** *Coherence in interactive multimedia*. Information needs

to be presented in a consistent manner with clear identification of the structure (Elges, 2003). Strategies include providing overview areas, maps, fixed display formats, and consistent placement of section titles.

2. **Clearly identify selectable areas:** *Navigation in interactive multimedia.*Clarity is accomplished by following standard Web conventions (i.e. underline and blue for active hyperlinks) or using icons that clearly indicate alternate navigation paths. A subprinciple to this guideline is to clearly indicate selections made, so that the users have a contextual understanding on where they have been and their current location.
3. **Indicate progress made:** *Interactivity in interactive multimedia.*This is an option particularly important when users are navigating through instructional material or taking an online assessment. Feedback on the status of the lecture or progression on the quiz eases navigation and favors cognition.
4. **Provide multiple versions of instructional material:** *Information Presentation*. This includes offering a text-only option, a text and graphics option, an audio narrated presentation, a video, or a variety of media accessible through high speed networks (Fleming, 1997). This is particularly important in order to guarantee broader accessibility (Johnson & Ruppert, 2002).
5. **Offer contextual help:** *Ease of use*.Contextual help facilitates navigation and ease of use (Tarafdar & Zhang, 2005). For example, if users experience difficulties in retrieving materials, specific browser options and configuration eases progress.
6. **Keep pages short:** *Screen design*.Scrolling may not be enjoyed by users (del Galdo & Nielsen, 1996). Information should be presented on sequential pages, providing the option to print the complete document through a single packaged file, conveniently placed in the first instructional screen.
7. **Link to other pages, not to other points in the same page:** *Mapping.*Long documents and text should be broken down in sequential pages. The users will have the ability to "jump" to other sessions or go back to the same paragraph by simply using back buttons and "breadcrumbs."
8. **Select links carefully:** *Cognitive load.*Too many links in the same page may overwhelm the student and disorient (del Galdo & Nielsen, 1996). Links should be placed only at the bottom of the page or at the end of the text that they refer to. Links conveniently placed within the paragraphs offer contextual information and clarification for the learners.
9. **Label links appropriately:** *Navigation*. Some textual links or icons may not clearly indicate the destination area. Particular attention needs to be paid to content synchronization.
10. **Keep important information at the top of the page:** *Screen Design.*As dynamic text, such as "flying" or moving effects, lower attention and focus (Yi & Davis, 2003), and are not supportive of learning, important information should be static and placed at the top of the page.
11. **Links and information must be kept updated:** *Information Presentation & Mapping.* Both content and links to other material need to be tested on a periodical basis to check the availability of the link ("active" links).
12. **Limit overly long download times:** *Interactivity.* As "traditional human factor guidelines indicate 10 seconds as the maximum response time before users lose interest" care should be used to decrease file size and download times (del Galdo & Nielsen, 1996; Tarafdar & Zhang, 2005).

In summary, the principles above map and extend guidelines applicable to interactive multimedia. These principles can provide guidance on how LMS supports learning by decreasing the cognitive load and increasing coherence. The next section presents an example on how these guidelines can be applied to uncover limitations with existing systems and systems under-development.

EXAMPLES OF EVALUATION: PROMETHEUS FORMATIVE EVALUATION

Having reviewed design principles and instructional objectives set forth in the literature, this section of the paper applies these principles to the evaluation of an online courseware application, Prometheus (Johnson & Ruppert, 2002). Prometheus was developed at the George Washington University starting in 1997. The Prometheus LMS evolved during its development through a series of formative evaluations (similar to the model presented in this section) and summative evaluations (described in the next section). The assessment process benefited the design of the system. It was conducted by a team of instructional designers at the Center for Instructional Design and Development of the university, including the author. The formative evaluations informed the development team of improvement needs. An example of the design guidelines is discussed below. The evaluation is conducted on a 5-point level (using Harvey balls to represent Very Low to Very High levels) and is summarized in Table 3.

Prometheus' main menu (navigation toolbox) identifies the structure of the course and the key course content. The main menu display is fixed, and coherently placed on each screen (see Figure 1).

Hyperlinks within pages are labeled with text descriptions and standard colors for visited/unvisited links are used. Additional hyperlinks that enable editing and interactivity are clearly identified by consistent yellow boxes placed at the top of each frame (see Figure 1).

Prometheus does not provide feedback on the progress made in the completion of the coursework. In the communication section of Prometheus, the discussion area, the provision of feedback on the navigation of the threaded/unthreaded messages is lacking. The user does not know how many messages are left to read when navigating sequentially through each discussion response. Figure 2 shows the navigation screen in the discussion area. Help on the contextual position of the user is missing (How many messages have been read? How many yet to read?).

Prometheus enables the integration of video and audio to any type of text and/or PowerPoint presentation. Faculty can deliver their lectures using a variety of media. A re-sizeable pop-up window with a multimedia presentation is available to students (see Figure 3). This window enables learner control (play, pause, and stop buttons) and self-paced learning.

Although Prometheus is a particularly user friendly interface, for example, no formal directions and Frequently Asked Questions (FAQ) responses to configure users' browsers are available. Contextual support with instructions on download could be easily integrated into Prometheus, rather than being handled individually and redundantly by each instructor.

The length of the pages in Prometheus vary depending on the amount of information each instructor uploads into the system. Prometheus pages may remain short, become very lengthy, depending on the instructor's preference for typing a lecture in Prometheus or simply uploading a document file that the students can download.

Prometheus does not enable hyperlinking within the same page (thus burdening the cognitive load). A new window will open if a file is being downloaded or a hyperlink has been selected.

Prometheus enables hyperlinking in specific and coherent areas. Although users can always

Table 3. Summary LMS evaluation

Web Design Guideline	Interactive Multimedia Equivalent	Prometheus
1. Provide structural clues	Coherence	●
2. Clearly identify selectable areas	Navigation	●
3. Indicate progress made	Interactivity	◔
4. Provide multiple version of instructional material	Information presentation	◕
5. Offering contextual help	Ease of use	○
6. Keep pages short	Screen design	◑
7. Link to other pages, not to other points in the same page	Mapping	●
8. Select links carefully	Cognitive load	◕
9. Label links appropriately	Navigation	◕
10. Keep important information at the top of the page	Screen design	◔
11. Links and information must be kept updated	Information presentation	◑
12. Limit overly long download times	Interactivity	◑

Legend: ○ Very Low ◔ Low ◑ Medium ◕ High ● Very High

Figure 1. Navigation and information presentation areas

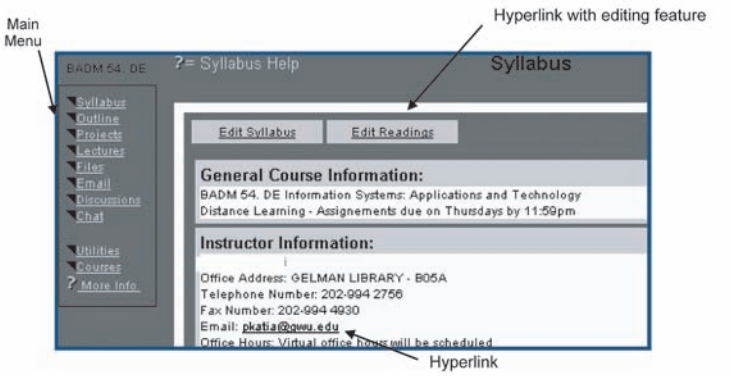


Figure 2. Feedback on navigation

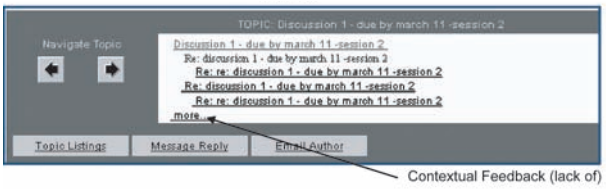
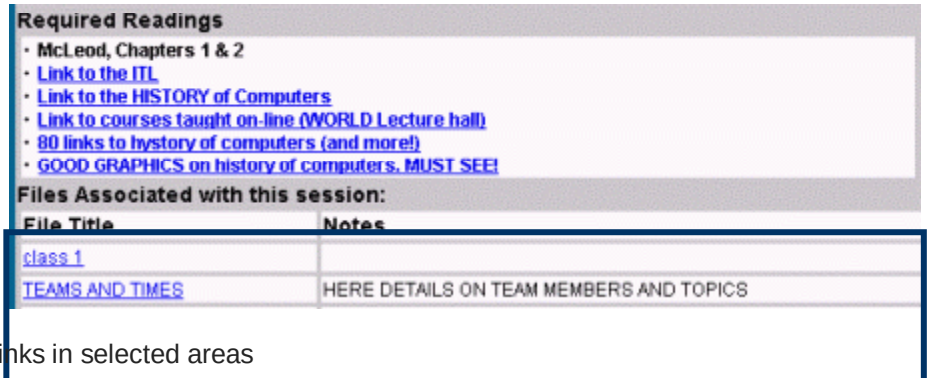


Figure 3. Multiple formats integration

Figure 4. Hyperlinks in selected areas



create links in any area by using HTML commands, Prometheus fill-in forms enable posting of URLs and other class materials only in selected areas (see "required reading for session x" or "files associated with session x" in Figure 4).

Links in Prometheus use text that is explanatory of the function that the selected area will accomplish. Contextual help that provides an overview of features is also available on selected screens. Labels are accurate.

Faculty has little control of the placement of information in Prometheus. Most of the placement in the interactive areas (files, projects, and discussion) is based on time of the posting and cannot be rearranged in different order. The forms in Prometheus can be used or left blank. If they are left blank, they do not appear on the screen. If filled, the order cannot be rearranged by the instructor by level of importance for the specific subject matter.

This criterion applies only to the information that is pertinent to the functioning of the system (and not the material posted by the instructor). A control mechanism to verify the "active" links and

restores back-up files is a needed improvement.

Although downloading time will vary depending on type of connection, modem speed and location (U.S. or abroad), the communication areas of Prometheus (i.e., discussion) suffer from long wait times to navigate through messages. Improvements in iterative releases of the software have decreased this problem, although it still remains substantial for users outside the campus.

Areas for improvement: the evaluation shows that Prometheus could be improved in:

- **Interactivity features:** Re-designing the discussion areas to provide contextual feedback and better navigation.
- **Technical support:** Offering users printable manuals and additional help on how to address the technical problems associated with browser configurations.
- **Screen customization (alias "spatial" and "temporal" synchronization):** Allow faculty and content developers to manipulate the layout and place the information that they consider most relevant in the top portions of the screen. A layout that constructs hierarchies of information based on the time of the posting is cumbersome. Allowing users to manipulate placement and order of uploaded information helps in the accomplishment of the learning objectives and guarantees that important information is not overlooked.

The implementation of the above recommendations, and iterative designs conducted from 1997 to 2003, enabled Prometheus to compete with commercial courseware applications and expanded its reach beyond the George Washington University community, for which it was originally intended. In 2003, Blackboard purchased the Prometheus system to integrate some of its developed features in their product offerings. A key factor in the decision to purchase the product was its high response to the needs of the teaching and learning community. This community was better served by using the results of the formative and summative evaluations described here. An evaluation guide for the assessment of LMS systems and their integration within a curriculum is included in the remainder of this paper to encourage an informed review of commercial applications. Suggestions for criteria and survey administration options are also included.

SUMMATIVE EVALUATION: ADMINISTRATION AND CRITERIA

While the formative evaluation was used as part of a process to improve the software during the development, it represented only a selected group of power users. Broader summative evaluations of the user population (faculty and students) enable corrective and developmental maintenance to comply with user expectations.

The evaluation instrument presented is developed on the basis of the interactive multimedia design principles earlier described. Each question in the scale (item development) is based on a set of related criteria for the evaluation of interactive multimedia products (Reeves, 1993). All the items in the scale are related to specific related domains. The survey measures attitudes and opinions on a self-reported 5-point Likert scale. Criteria for evaluation are based on the perception of interface design and the perception of usefulness of the application by users (students and faculty).

Administration

The survey questions (see Appendix 1) should be administered to two groups of users (faculty and students). Timing of the survey administration is an important factor — it should take place preferably at the end of an academic semester. In order to enable the evaluation of features that were used in the classroom, respondents should be enabled to access only questions relative to the features they used.

Participation in the survey questionnaire may vary. Different strategies could be used to encourage all system users to complete the online survey. For example, incentives could be offered, such as a drawing of free computer software could be conducted for all student respondents. Similarly, faculty participation could be encouraged. Alternatively, participation in the survey could be required of all users (as long as anonymity is guaranteed). For example, users may not be able to access any of the features before they complete the online questionnaire. To avoid user frustration or disruption of user's work schedule, users could be warned that they will be able to access only 10 additional working sessions, before the system will prompt them to complete the survey in order to be able to proceed. They may choose to take the survey earlier, but they should be informed and given enough time to complete important tasks, before the system locks them out.

Both approaches have pros and cons. A survey that is completely voluntary may not get enough responses, or may suffer from a response bias. A compulsory survey may frustrate some users, but will engage the entire user population.

Summative Evaluation Criteria

Reeves' evaluation criteria (1993) focus on the user interface of interactive instructional products, such as multimedia programs. As mentioned earlier, these criteria extend to LMS. If the user interface is not well designed, users will have little opportunity to learn from the program. Examples of a student survey instrument are included in Appendix 1 and key criteria used to define the survey questions (based on Reeves, 1993) are presented in this section. Continuing on the description of the earlier example, the sample questions are referred to a specific LMS system (Prometheus).

Ease of Use

"Ease of Use" is concerned with the perceived ease of a user interaction with the program. Figure 5 illustrates Reeves' dimension as ranging from the perception that the program is very difficult to use to one that is perceived as being very easy to use.

Navigation

"Navigation" is concerned with the perceived ability to move through the contents of an interactive program in an intentional manner. Figure 6 illustrates Reeves' dimension of interactive multimedia ranging from the perception that a program is difficult to navigate to one that is perceived as being easy to navigate. Possible options for navigation include evaluating the clarity of navigation icons.

Figure 5. "Ease of Use"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
I was able to learn Prometheus on my own ……..	❍	❍	❍	❍	❍	❍
Prometheus menus are intuitive……..	❍	❍	❍	❍	❍	❍

Figure 6. "Navigation"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The navigation options in Prometheus are clear in each section……	❍	❍	❍	❍	❍	❍

Figure 7. "Cognitive Load"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
I do not need to remember several commands to use Prometheus……..……	❍	❍	❍	❍	❍	❍

Figure 8. "Mapping"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
Prometheus navigation layout is consistent……..….	❍	❍	❍	❍	❍	❍

Cognitive Load

The cognitive load is defined as any effort in addition to reading affects comprehension (i.e., navigation efforts or adjustment to the user interface). The higher the cognitive load the more difficult it is for the learner to comprehend. In terms of "cognitive load", Reeves states that the user interfaces can seem unmanageable (i.e., confusing) or easily manageable (see Figure 7).

Mapping

"Mapping" refers to the program's ability to track and graphically represent to the user the navigation path through the program. This is a critical variable because users frequently complain of being lost in an interactive program. Evaluations of interactive programs vary from containing no mapping function to an appropriately powerful mapping function (see Figure 8).

Screen Design

Screen design is a dimension of interactive programs that evaluates elements such as text (font layout and type), icons, graphics (placement), color (balance), and other visual aspects of interactive programs. "Screen design" ranges from substantial violations of the principles of screen design to general adherence of these principles (see Figure 9).

Knowledge Space Compatibility — [Content]

Refers to the compatibility of the product content with the layout of the learning space in the software application. When a novice user initiates a search for information in an interactive program, s/he could perceive the resulting information as compatible with his/her current knowledge space (see Figure 10). If the search results are not compatible, the application is weak in integrating content and technical features. This criterion is mostly applicable in interactive multimedia, where the content placement is static. In LMS systems, it can be used as "content" evaluation instruments.

Information Presentation

A dimension concerned with whether the information contained in an interactive program is presented in an understandable form. A well-designed user interface is ineffective if the information it is intended to present is incomprehensible to the user. (see Figure 11).

Media Integration

Deals with the question on whether the various media (text, graphics, audio, video) work together to form one cohesive program. The media integration dimension is defined as ranging from uncoordinated to coordinated (see Figure 12).

This criterion is not applicable in the context of LMS because the integration of media and their

Figure 9. "Screen Design"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The text layout on the screen makes it easy to read……..	❍	❍	❍	❍	❍	❍

Figure 10. "Knowledge Space Compatibility"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
I can understand the meaning of all the instructions on any Prometheus page……..	❍	❍	❍	❍	❍	❍

Figure 11. "Information Presentation"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
Prometheus enables me to access class material in an organized way……..	❍	❍	❍	❍	❍	❍

Figure 12. "Media Integration"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
Prometheus enables me to easily interact with my instructor……	❍	❍	❍	❍	❍	❍

Figure 13. "Aesthetics"

Example on a 5-point Likert scale:

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
Prometheus screen design is pleasing…….	❍	❍	❍	❍	❍	❍

quality will be dependent primarily on the quality of the application that the individual content developers (faculty) will upload in the courseware. This criterion can be substituted with questions relative to "class interaction" and collaboration tools, key components of LMS tools that support interaction in multiple ways (audio, voice, and text interaction).

Aesthetics

"Aesthetics" deals with a subjective evaluation of the user of the screen layout ranging from displeasing to pleasing (see Figure 13).

Overall Functionality

"Overall Functionality" is related to the perceived utility of the program to achieve what its intended purposes are. It will include an evaluation of the factors that affect the perceived quality of the application. Figure 14 illustrates a dimension of the user interface of interactive programs that ranges from dysfunctional to highly functional.

Additional Criteria

Since Prometheus contains a series of features (course design tools and collaboration tools) that

Figure 14. "Overall Functionality"

Example on a 5-point

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The technical quality of Prometheus is satisfactory..…	❍	❍	❍	❍	❍	❍

enable different types of class interaction, evaluation of the usefulness of the individual feature (as perceived by the user) is an important component of a complete summative evaluation. Questions evaluating user's perception of usefulness of the system will vary depending on whether the user is a student or a faculty member. The questions will cover each of the features available to the users, but will be accessed by the user only if s/he reported being familiar with or having used the feature at the beginning of the survey (see in Appendix 1, Syllabus, Outline, Projects, Lectures, Files, Email, Discussion, Chat, Utilities questions).

The deployment of the survey to a large population of LMS users can provide a better understanding of how LMS supports learning.

CONCLUSION

This paper presents a framework for the evaluation of LMS based on the criteria set forth by the literature on interactive multimedia. It claims that the convergence between multimedia and Web-based learning environments enables the extension of design guidelines to LMS. As in interactive instructional multimedia systems, an effective LMS strives for coherence and focuses on the reduction of the learner's cognitive load. LMS systems are currently consolidating, but variations and customizations still exist, especially in emerging open-source products representing lower cost solutions (Hall, 2005).

Evaluating LMS systems remains a key prerequisite and a first step for evaluating Web-based instruction effectiveness (Hills, 2003a). To address this assessment need, an evaluation protocol for LMS was proposed in this paper based on the integration of interactive multimedia and Web-evaluation criteria. Different types of evaluation and evaluation criteria were presented. Sample questions for each evaluation and strategies for the survey administration were also briefly discussed. These questions may constitute a useful reference tool for LMS evaluations. The survey questionnaire presented in the Appendix could be reviewed, changed, and integrated beyond the issues discussed in this paper. In any event, it could serve as a starting point for a serious effort to evaluate instructional software that has not yet been assessed by the majority of its users before, during, or after its integration in the curriculum. As Stoner (1996) and other authors (Carmean & Haefner, 2002) point out, *evaluation* is a key element of any proper curriculum implementation.

REFERENCES

Alessi, S., & Trollip, S. (1991). *Computer-based instruction: Methods and development.* Englewood Cliffs, NJ: Prentice Hall.

Bersin, J. (2005). Evaluating LMSs? Buyer beware. *Training, 42,* 26-31.

Carmean, C., & Haefner, J. (2002). Mind over matter: Transforming course management systems into effective learning environments. *EDUCAUSE Review, 37,* 26-34.

Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary Education and Management, 11,* 19-36.

del Galdo, E. M. & Nielsen, J. (Eds.). (1996). *International user interfaces.* New York: John Wiley & Sons.

Draper, S., Henderson, F., Brown, M., & McAteer, E., (1996). Integrative evaluation: An emerging role for classroom studies of CAL. *Computers and Education, 26*(1-3), 17-32.

Elges, M. (2003). Designing for Web accessibility: More benefits than you may imagine. *Nonprofit World, 21,* 26-28.

Fleming, D. (1997). Dynamite Webpage design. *Training & Development, 51,* 51-52.

Gagné, R., Briggs, L., & Wagner, W. (1988). *Principles of instructional design* (3rd ed.). New York:Holt Reinbank.

Gurak, L. (1992). Towards consistency in visual information: Standardized icons based on task. *Technical Communication*, (First Quarter), 33-37.

Hall, B. (2005). Low-cost-LMSs. *Training, 42,* 36.

Hall, S., &. Hall (2004). A guide to learning content management systems. *Training, 41,* 33-37.

Hedberg, J. Brown, C., & Arrighi, M. (1997). Interactive multimedia and Web-based learning: Similarities and Differences. In B. Kahn (Ed), *Web-based instruction.* Englewood Cliffs, NJ: Educational Technology Publications.

Hills, H. (2003a). Learning management systems Part 2: The benefits they can promise. *Training Journal, 20,* February.

Hills, H. (2003b). Learning management systems Part 3: Making the right decisions. *Training Journal*, 34, March.

Hills, H. (2003c). Learning management systems: Why buy one? *Training Journal*, 12, January.

Johnson, A., & Ruppert, S. (2002). An evaluation of accessibility in online learning management systems. *Library Hi Tech, 20,* 441-451.

Jones, M., Farquhar, J., & Surry, D. (1995). Using meta-cognitive theories to design user interfaces for computer-based learning. *Educational Technology, 35*(4), 12-22.

Lucas, H. (1994). *Information systems concepts for management* (5th ed.). New York: McGraw-Hill.

Milheim, W. D., & Lavix, C. (1992). Screen design for computer-based training and interactive video: Practical suggestions and overall guidelines. *Performance and Instruction, 31*(5), 13–21.

Mukherjee, P., & Edmonds, G. (1993). Screen design: A review of research. (ERIC Document Reproduction Service No. ED 370 561).

O'Brien, J. (2005). *Introduction to information systems* (12th ed.). New York: McGraw-Hill.

Orr, K., Golas, K., & Yao, K. (1994, Winter). Storyboard development for interactive multimedia training. *Journal of Interactive Instruction Development,* pp. 18-31.

Overbaugh, R. (1994). Research-based guidelines for computer-based instruction development. *Journal of Research on Computing in Education, 27*(1), 29-47.

Reeves, T. (1993). Evaluating interactive multimedia. In D. Gayesky (Ed.), *Multimedia for learning, development, application, evaluation.*

(pp. 97-112). Englewood Cliffs, NJ: Educational Technology.

Stemler, K. (1997). Educational characteristics of multimedia: A literature review. *Journal of Educational Multimedia and Hypermedia*, *6*(3,4).

Stoner, G. (1996). *Implementing learning technology*. Learning Technology Dissemination Initiative. Retrieved November 10, 2005, from http://www.icbl.hw.ac.uk/ltdi/implementing-it/cont.htm

Szabo, M., & Kanuka, H. (1999). Effects of violating screen design principles of balance, unity, and focus on recall learning, study time, and completion rates. *Journal of Educational Multimedia and Hypermedia*, *8*(1), 23-42.

Tarafdar, M., & Zhang, J. (2005). Analyzing the influence of Web site design parameters on Web site usability. *Information Resources Management Journal*, *18*(4), 62-80.

Thuring, M., J. Hannemann, & J. Haake (1995). Hypermedia and cognition: Designing for comprehension. *Communications of the ACM 38*(8), 57-66.

Taylor, P. (2003, June 23). Market in fresh mood of realism: Learning management systems — Customers face difficult choices from a big range of systems. *Financial Times.*

Watson, J., & Hardaker G. (2005). Steps towards personalized learner management system (LMS): SCORM implementation. *Campus-Wide Information Systems, 22*, 56-70.

Yi, M., & Davis, F. (2003). Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research*, *14*, 146.

APPENDIX 1. SUMMATIVE EVALUATION QUESTIONNAIRE

Sample Student Survey

Please indicate how many courses you took on the LMS
_____(number)

Please indicate which features of the LMS your courses used (check all that apply)

- ❑ Syllabus
- ❑ Projects
- ❑ Lectures
- ❑ Files

- ❑ Email
- ❑ Discussion
- ❑ Chat

Please evaluate the LMS based on your overall experience with this Web-based courseware and your experience with the individual features used. Remember that this is an evaluation of the LMS as a software application, and **not** an evaluation of how well your instructor used the LMS.

Thank you for your time!

[Ease of Use]

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The LMS menus are intuitive……..…….	❍	❍	❍	❍	❍	❍
If I have any problem, The LMS help menu provides useful information…..	❍	❍	❍	❍	❍	❍
I was able to learn The LMS features on my own……..	❍	❍	❍	❍	❍	❍
If I needed help, I had plenty of opportunities to learn additional The LMS features through:						
- The LMS team support ……...	❍	❍	❍	❍	❍	❍
- Support from my instructor ……	❍	❍	❍	❍	❍	❍
- Support from other students……..	❍	❍	❍	❍	❍	❍

Which other sections would be useful to you?

1. _____
2. _____
3. _____

[Screen Design]

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The text layout on the screen makes it easy to read…..……..	❍	❍	❍	❍	❍	❍

[Content]

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
I can understand the meaning of all the instructions on any of the LMS page…….	❍	❍	❍	❍	❍	❍

Which instructions are unclear?

1. _____
2. _____
3. _____

[Information Presentation]

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The LMS enables me to access class material in an organized way…….……	❍	❍	❍	❍	❍	❍

Please, indicate which type of documents you accessed on The LMS (check all that applies).

- ❑ Plain Text
- ❑ Word documents
- ❑ HTML files
- ❑ Video files
- ❑ Sound files
- ❑ PowerPoint presentations
- ❑ Portable document format (.PDF)
- ❑ Steaming media files (narrated PPT, audio, video)

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The media (video, sound, graphics or text) used by my instructors on The LMS enabled me to better understand class information……	❍	❍	❍	❍	❍	❍

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree	Not Applicable
The LMS is a useful supplement in my courses…….……...	❍	❍	❍	❍	❍	❍

Which sections of The LMS gave you the most technical problems when accessing them?

1. _____
2. _____
3. _____

Which type of problems did you have?

1. _____
2. _____
3. _____

Have you used any other LMS?

- ❑ Yes (please specify_____)
- ❑ No

If yes, how does it compare with The LMS?

- ❑ Better
- ❑ The same
- ❑ Not as good

Section Specific Questions

[note: Questions in this section should appear in an online survey only if respondent selected that he/she uses the specific The LMS feature]

This work was previously published in Online Distance Learning: Concepts, Methodologies, Tools, and Applications, edited by L. Tomei, pp. 805-821, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.8
Multimedia Interactivity on the Internet

Omar El-Gayar
Dakota State University, USA

Kuanchin Chen
Western Michigan University, USA

Kanchana Tandekar
Dakota State University, USA

INTRODUCTION

With the interactive capabilities on the Internet, business activities such as product display, order placing and payment are given a new facelift (Liu & Shrum, 2002). Consumer experience is also enhanced in an interactive environment (Haseman, Nuipolatoglu & Ramamurthy, 2002). A higher level of interactivity increases the perceived telepresence and the user's attitude towards a Web site (Coyle & Thorson, 2001). When it comes to learning, a higher level of interactivity improves learning and learner satisfaction (Liu & Schrum, 2002). While interactivity does not necessarily enable enhanced gain in user learning, it positively influences learners' attitudes (Haseman et al., 2002). Interactivity has been shown to engage users in multimedia systems (Dysart, 1998) to encourage revisits to a Web site (Dholakia et al., 2000), to increase satisfaction toward such systems (Rafaeli &Sudweeks, 1997), to enhance the visibility (as measured in number of referrals or backward links) of Web sites (Chen & Sockel, 2001) and to increase acceptance (Coupey, 1996).

BACKGROUND

According to the Merriam Webster dictionary, "interactivity" refers to 1) being mutually or reciprocally active, or 2) allowing two-way electronic communications (as between a person and a computer). However, within the scientific community, there is little consensus of what interactivity is, and the concept often means different things to different people (Dholakia, Zhao, Dholakia & Fortin, 2000; McMillan & Hwang,

2002). McMillan and Hwang (2002) suggest that interactivity can be conceptualized as a process, a set of features and user perception. Interactivity as a process focuses on activities such as interchange and responsiveness. Interactive features are made possible through the characteristics of multimedia systems. However, the most important aspect of interactivity lies in user perception of or experience with interactive features. Such an experience may very likely be a strong basis for future use intention.

Interactivity is considered a process-related construct, where communication messages in a sequence relate to each other (Rafaeli & Sudweeks, 1997). Ha and James (1998, p. 461) defined interactivity as "the extent to which the communicator and the audience respond to, or are willing to facilitate, each other's communication needs." Interactions between humans via media are also called mediated human interactions or computer-mediated communication (Heeter, 2000). Early studies tend to consider interactivity as a single construct, where multimedia systems vary in degrees of interactivity. Recent studies suggest that interactivity is a multi-dimensional construct.

As research continues to uncover the dynamic capabilities of multimedia systems, the definition of interactivity evolves to include aspects of hardware/software, processes during which the interactive features are used and user experience with interactive systems. Dholakia et al. (2000) suggest the following six interactivity dimensions: 1) user control, 2) responsiveness, 3) real-time interactions, 4) connectedness, 5) personalization/customization, and 6) playfulness. Similarly, Ha and James (1998) suggest five interactivity dimensions: 1) playfulness, 2) choice, 3) connectedness, 4) information collection, and 5) reciprocal communication.

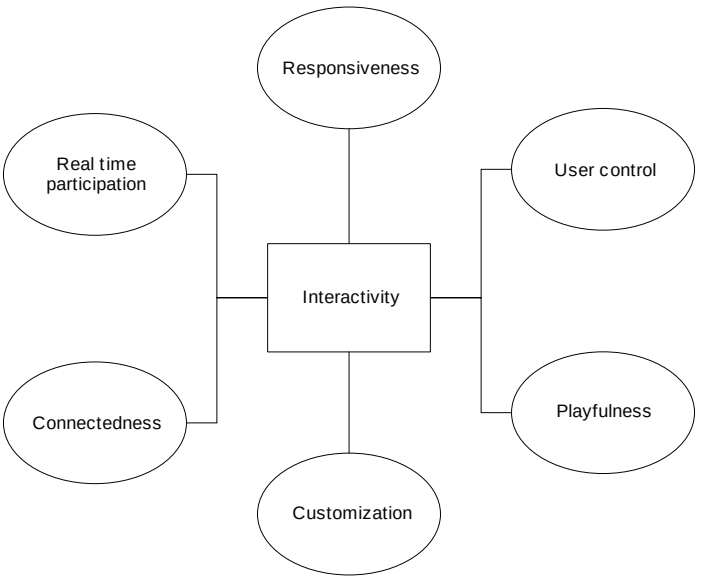
Within the context of multimedia systems, we view interactivity as a multidimensional concept referring to the nature of person-machine interaction, where the machine refers to a multimedia system. Figure 1 presents a conceptual framework, including interactivity dimensions defined as follows:

- **User control:** The extent to which an individual can choose the timing, content and sequence of communication with the system.
- **Responsiveness:** The relatedness of a response to earlier messages (Rafaeli & Sudweeks, 1997).
- **Real-time participation:** The speed with which communication takes place. This can range from instant communication (synchronous) to delayed response communication (asynchronous).
- **Connectedness:** The degree to which a user feels connected to the outside world through the multimedia system (Ha & James, 1998).
- **Personalization/Customization:** The degree to which information is tailored to meet the needs of individual users. For example, interactive multimedia learning systems must be able to accommodate different learning styles and capabilities.
- **Playfulness:** The entertainment value of the system; that is, entertainment value provided by interactive games or systems with entertaining features.

TECHNOLOGIES AND PRACTICES

The ubiquity of multimedia interactivity in general and on the Internet in particular is realized through the exponential growth in information technology. Specifically, the growth in computational power enabling ever-increasingly multimedia features coupled with advances in communication technologies and the Internet are pushing the interactivity frontier. Such technologies include, but are not limited to, a range of technologies, from the basic point and click to highly complex multimedia systems.

Figure 1. Interactivity as a multidimensional concept



In practice, and in their quest for interactivity, companies and organizations have resorted to a variety of techniques to encourage interactions in their systems. Table 1 provides a framework to map important multimedia/Web features from the existing literature to the six interactivity dimensions discussed in Figure 1. The goal of this framework is to offer practitioners a basis to evaluate interactivity in their multimedia systems. For example, a Web site designer may want to compare his or her design with popular Web sites in the same industry to measure if they offer a similar level of interactivity. Two important issues concerning the comparison include what interactive features are recommended for comparison and how to quantify interactivity features for comparison. The framework in Table 1 serves to answer the first question. One way to answer the second question involves simply counting the number of interactivity features in each of the interactivity dimensions. This counting technique is referred to as the interactivity index (II) and is frequently used by researchers to quantify interactivity. The quantified results, if measured consistently, can be used for longitudinal or cross-industry comparisons. Additionally, interactivity is examined with other constructs. Readers interested in empirical results focusing on the relationship between interactivity dimensions and other constructs are referred to the cited references, such as Ha and James (1998), Dholakia et al. (2000); Chen and Sockel (2001); McMillan and Hwang (2002); Burgoon, Bonito, Ramirez, Dunbar, Kam and Fischer (2002); and Chen and Yen (2004).

CURRENT RESEARCH

Interactivity is an active area of research that spans a number of research fields, including computer science, human computer interaction (HCI), information systems, education, marketing, advertisement and communication. A comprehensive review of the literature is beyond the scope of this article. Instead, we focus our attention on current research effort as it pertains to multimedia interactivity on the Internet, with a particular emphasis on education, advertisement and marketing.

Table 1. A framework of mapping multimedia/Web features to interactivity dimensions

Interactivity dimensions	Multimedia/Web features	
User control	• Alternative options for site navigation • Linear interactivity, where the user is able to move (forward or backwards) through a sequence of contents	• Object interactivity (proactive inquiry) where objects (buttons, people or things) are activated by using a pointing device.
Responsiveness	• Context-sensitive help • Search engine within the site	• Dynamic Q&A (questions and responses adapt to user inputs)
Real-time participation	• Chat rooms • Video conferencing	• E-mail • Toll-free number
Connectedness	• Video clips • Site tour	• Audio clips • Product demonstration
Personalization/ Customization	• Site customization • Bilingual site design	• Customization to accommodate browser differences
Playfulness	• Games • Software downloads • Visual simulation	• Online Q&A • Browser plug-ins (e.g., flash, macromedia, etc.)

Current research on multimedia interactivity predominantly focuses on conceptual issues related to the definition and measurement of interactivity, evaluation of interactive multimedia systems, design issues and applications of interactive multimedia systems. Regarding conceptual issues, Kirch (1997) questions the decision cycle model, which is the received theory in human computer interaction, and discusses additional ways of interacting with multimedia systems; while Ohl (2001) questions the adequacy of current definitions of interactivity in the context of educational systems.

Haseman, Polatoglu and Ramamurthy (2002) found that interactivity leads to favorable attitude formation but not so much to improved learning outcomes. There had been no evidence to prove that interactivity influences user achievement. Liu and Shrum (2002) propose that higher levels of interactivity create a cognitively involving experience and can enhance user satisfaction and learning.

Concerning design considerations, Robinson (2004) identifies interactivity as one of eight principles for the design of multimedia material. Examples of case studies and applications reported in the literature include Abidin and Razak's (2003) presentation of Malay folklore using interactive multimedia. Table 2 lists research contributions pertaining primarily to multimedia interactivity.

Internet interactivity has also attracted interest in areas such as the measurement of interactivity, evaluation of the effectiveness of interactivity and design considerations for Internet-interactive Web sites. For example, Paul (2001) analyzed the content of 64 disaster relief Web sites and found that most sites had a moderate level of interactivity but were not very responsive to their users. A study conducted by Ha and James (1998) attempted to deconstruct the meaning of interactivity, and then reported the results of a content analysis that examined the interactivity levels of business Web sites. Their findings suggest that five interactivity dimensions are possible, with the reciprocal

Table 2. Current research focusing primarily on multimedia interactivity

Research focus	Conceptual	Evaluation	Design	Application
Research work	Ohl (2001), Massey (2000)	Karayanni et al. (2003), Haseman et al. (2002), Liu and Shrum (2002), Ellis (2001), Moreno (2001), Mayer (2001)	Robinson (2004), Zhang et al. (2003), Trindade et al. (2002)	Adibin et al. (2003), Hou et al. (2002), Paustian (2001)

communication dimension being the most popular dimension. In an effort to explore the relationship between Ha and James' (1998) interactivity dimensions and the quality of Web sites, Chen and Yen (2004) suggested that reciprocal communication, connectedness and playfulness are the most salient dimensions of interactivity that influence design quality. Moreover, Lin and Jeffres (2001) performed a content analysis of 422 Web sites associated with local newspapers, radio stations and television stations in 25 of the largest metro markets in the United States. Results show that each medium has a relatively distinctive content emphasis, while each attempts to utilize its Web site to maximize institutional goals.

According to Burgoon et al. (2002), computer mediated communication may even be better than non-mediated or face-to-face interaction, even though face-to-face is considered easier. The study also points out that distal communication, mediation and loss of non-verbal cues do not necessarily result in worse decision quality or influence, but may, in fact, enhance performance in some cases.

Addressing design considerations, McMillan (2000) identified 13 desirable features that an interactive Web site should possess in order to be interactive. These features include: e-mail links, hyperlinks, registration forms, survey forms, chat rooms, bulletin boards, search engines, games, banners, pop-up ads, frames and so forth. High levels of vividness help create more enduring attitudes (Coyle & Thorson, 2001). A study by Bucy, Lang, Potter and Grabe (1999) found presence of advertising in more than half of the Web pages sampled. It also suggests a possible relationship between Web site traffic and the amount of asynchronous interactive elements like text links, picture links, e-mail links, survey forms and so forth. Features most commonly used on the surveyed Web sites were frames, logos and a white background color.

FUTURE TRENDS

Long-term impacts of interactivity should be studied on learning, attitudes and user outcomes. To study learning behavior of students, their knowledge should be tested twice; once at first and then after a few days or weeks for absorption/retention (Haseman et al., 2002). Coyle and Thorson (2001, p. 76) suggested to "focus on additional validation of how new media can approximate a more real experience than traditional media." One way to do this would be to replicate previous findings dealing with direct or indirect experience. Will more interactive and more vivid systems provide more direct experience than less interactive, less vivid systems? Also, future research should focus on testing specific tools to understand how their interactivity characteristics improve or degrade the quality of user tasks at hand.

The current literature appears to lack consensus on the dimensionality of interactivity. Inconsistent labeling or defining the scope of interactivity dimensions exists in several studies; for example, playfulness and connectedness ap-

pear to be included in both Dholakia et al. (2000) and Ha and James (1998), but Dholakia et al.'s personalization/customization dimension was embedded in Ha and James' choice dimension. Furthermore, much of interactivity research employed only qualitative assessment of interactivity dimensions (such as Heeter, 2000), suggesting future avenues for empirical validations and perhaps further refinement.

Despite disagreements in interactivity dimensions, user interactivity needs may vary across time, user characteristics, use contexts and peer influence. A suggestion for further research is to take into account the factors that drive or influence interactivity needs in different use contexts. Another suggestion is to study whether user perception depends on the emotional, mental and physical state of people; that is, their personality and to what extent or degree it depends on these characteristics and how these can be altered to improve the overall user perception.

CONCLUSION

Multimedia interactivity on the Internet – while considered as "hype" by some – is here to stay. Recent technological advancements in hardware, software and networks have enabled the development of highly interactive multimedia systems. Studying interactivity and its effects on target users certainly impact business values. Research pertaining to interactivity spans a number of disciplines, including computer science, information science, education, communication, marketing and advertisement. Such research addressed a variety of issues, ranging from attempting to define and quantify interactivity to evaluating interactive multimedia systems in various application domains, to designing such systems. Nevertheless, a number of research issues warrant further consideration, particularly as it pertains to quantifying and evaluating interactive multimedia systems.

In effect, the critical issues discussed in this chapter offer many implications to businesses, governments and educational institutions. With regard to businesses, multimedia interactive systems will continue to play a major role in marketing and advertisement. Interactive virtual real estate tours are already impacting the real estate industries. Interactive multimedia instruction is changing the way companies and universities alike provide educational services to their constituents. From physics and engineering to biology and history, interactive multimedia systems are re-shaping education.

REFERENCES

Abidin, M.I.Z., & Razak, A.A. (2003). Malay digital folklore: Using multimedia to educate children through storytelling. *Information Technology in Childhood Education Annual*, (1), 29-44.

Bucy, E.P., Lang, A., Potter, R.F., & Grabe, M.E. (1999). Formal features of cyberspace: Relationship between Web page complexity and site traffic. *Journal of the American Society for Information Science*, *50*(13), 1246-1256.

Burgoon, J.K., Bonito, J.A., Ramirez, A., Dunbar, N.E., Kam, K., & Fischer, J. (2002). Testing the interactivity principle: Effects of mediation, propinquity, and verbal and nonverbal modalities in interpersonal interaction. *Journal of Communication*, *52*(3), 657-677.

Chen, K., & Sockel, H. (2001, August 3-5). Enhancing visibility of business Web sites: A study of cyber-interactivity. *Proceedings of Americas Conference on Information Systems*, (pp. 547-552).

Chen, K., & Yen, D.C. (2004). Improving the quality of online presence through interactivity. *Information & Management*, forthcoming.

Coupey, E. (1996). Advertising in an interactive environment: A research agenda. In D.W. Schumann & E. Thorson (Eds.), *Advertising and the World Wide Web* (pp. 197-215). Mahwah, NJ: Lawrence Erlbaum Associates.

Coyle, J.R., & Thorson, E. (2001). The effects of progressive levels of interactivity and vividness in Web marketing sites. *Journal of Advertising, 30*(3), 65-77.

Dholakia, R.R., Zhao, M., Dholakia, N., & Fortin, D.R. (2000). Interactivity and revisits to Web sites: A theoretical framework. Research institute for telecommunications and marketing. Retrieved from *http://ritim.cba.uri.edu/wp2001/wpdone3/Interactivity.pdf*

Dysart, J. (1998). Interactivity: The Web's new standard. *NetWorker: The Craft of Network Computing, 2*(5), 30-37.

Ellis, T.J. (2001). Multimedia enhanced educational products as a tool to promote critical thinking in adult students. *Journal of Educational Multimedia and Hypermedia, 10*(2), 107-124.

Ha, L. (2002, April 5-8). Making viewers happy while making money for the networks: A comparison of the usability, enhanced TV and TV commerce features between broadcast and cable network Web sites. *Broadcast Education Association Annual Conference*, Las Vegas, Nevada.

Ha, L., & James, E.L. (1998). Interactivity reexamined: A baseline analysis of early business Web sites. *Journal of Broadcasting & Electronic Media, 42*(4), 457-474.

Haseman, W.D., Nuipolatoglu, V., & Ramamurthy, K. (2002). An empirical investigation of the influences of the degree of interactivity on user-outcomes in a multimedia environment. *Information Resources Management Journal, 15*(2), 31-41.

Heeter, C. (2000). Interactivity in the context of designed experiences. *Journal of Interactive Advertising, 1*(1). Available at *www.jiad.org/vol1/no1/heeter/index.html*

Hou, T., Yang, C., & Chen, K. (2002). Optimizing controllability of an interactive videoconferencing system with Web-based control interfaces. *The Journal of Systems and Software, 62*(2), 97-109.

Karayanni, D.A., & Baltas, G.A. (2003). Web site characteristics and business performance: Some evidence from international business-to-business organizations. *Marketing Intelligence & Planning, 21*(2), 105-114.

Lin, C.A., & Jeffres, L.W. (2001). Comparing distinctions and similarities across Web sites of newspapers, radio stations, and television stations. *Journalism and Mass Communication Quarterly, 78*(3), 555-573.

Liu, Y., & Shrum, L.J. (2002). What is interactivity and is it always such a good thing? Implications of definition, person, and situation for the influence of interactivity on advertising effectiveness. *Journal of Advertising, 31*(4), 53-64.

Massey, B.L. (2000). Market-based predictors of interactivity at Southeast Asian online newspapers. *Internet Research, 10*(3), 227-237.

Mayer, R.E., & Chandler, P. (2001). When learning is just a click away: does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*(2), 390-397.

McMillan, S.J. (2000). Interactivity is in the eye of the beholder: Function, perception, involvement, and attitude toward the Web site. In M.A. Shaver (Ed.), *Proceedings of the American Academy of Advertising* (pp. 71-78). East Lansing: Michigan State University.

McMillan, S. J., & Hwang, J. (2002). Measures of perceived interactivity: An exploration of the role of direction of communication, user control, and time in shaping perceptions of interactivity. *Journal of Advertising, 31*(3), 29-42.

Moreno, R., Mayer, R.E., Spires, H., & Lester, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*(2), 177-213.

Ohl, T.M. (2001). An interaction-centric learning model. *Journal of Educational Multimedia and Hypermedia*, *10*(4), 311-332.

Paul, M.J. (2001). Interactive disaster communication on the Internet: A content analysis of 64 disaster relief. *Journalism and Mass Communication Quarterly*, *78*(4), 739-753.

Paustian, C. (2001). Better products through virtual customers. *MIT Sloan Management Review*, *42*(3), 14.

Rafaeli, S., & Sudweeks, F. (1997). Networked interactivity. *Journal of Computer-Mediated Communication*, *2*(4). Available at *www.ascusc.org/jcmc/vol2/issue4/rafaeli.sudweeks.html*

Robinson, W.R. (2004). Cognitive theory and the design of multimedia instruction. *Journal of Chemical Education*, *81*(1), 10.

Trindade, J., Fiolhais, C., & Almeida, L. (2002). Science learning in virtual environments: a descriptive study. *British Journal of Educational Technology*, *33*(4), 471-488.

Zhang, D., & Zhou, L. (2003). Enhancing e-learning with interactive multimedia. *Information Resources Management Journal*, *16*(4), 1-14.

KEY TERMS

Computer-Mediated Communication (CMC): Refers to the communication that takes place between two entities through a computer, as opposed to face-to-face interaction that takes place between two persons present at the same time in the same place. The two communicating entities in CMC may or may not be present simultaneously.

Machine Interactivity: Interactivity resulted from human-to-machine or machine-to-machine communications. Typically, the later form is of less interest to most human-computer studies.

Reach: To get users to visit a Web site for the first time. It can be measured in terms of unique visitors to a Web site.

Reciprocal Communication: Communication that involves two or more (human or non-human) participants. The direction of communication may be two way or more. However, this type of communication does not necessarily suggest that participants communicate in any preset order.

Stickiness: To make people stay at a particular Web site. It can be measured by time spent by the user per visit.

Synchronicity: It refers to the spontaneity of feedback received by a user in the communication process. The faster the received response, the more synchronous is the communication.

Telepresence: Defined as the feeling of being fully present at a remote location from one's own physical location. Telepresence creates a virtual or simulated environment of the real experience.

Two-Way Communication: Communication involving two participants; either both of the participants can be humans or it could be a human-machine interaction. It does not necessarily take into account previous messages.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 724-730, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.9
Multimedia Content Adaptation

David Knight
Brunel University, UK

Marios C. Angelides
Brunel University, UK

INTRODUCTION

The previous decade has witnessed a wealth of advancements and trends in the field of communications and subsequently, multimedia access. Four main developments from the last few years have opened up the prospect for ubiquitous multimedia consumption: wireless communications and mobility, standardised multimedia content, interactive versus passive consumption and the Internet and the World Wide Web. While individual and isolated developments have produced modest boosts to this existing state of affairs, their combination and cross-fertilisation have resulted in today's complex but exciting landscape. In particular, we are beginning to see delivery of all types of data for all types of users in all types of conditions (Pereira & Burnett, 2003).

Compression, transport, and multimedia description are examples of individual technologies that are improving all the time. However, the lack of interoperable solutions across these spaces is holding back the deployment of advanced multimedia packaging and distribution applications. To enable transparent access to multimedia content, it is essential to have available not only the description of the content but also a description of its format and of the usage environment in order that content adaptation may be performed to provide the end-user with the best content experience for the content requested with the conditions available (Vetro, 2003).

In the following sections, we will look at the background of multimedia content adaptation, why do we require it and why are present solutions not adequate. We then go onto the main focus of the article, which describes the main themes of modern multimedia content adaptation, such as present day work that defines the area and overviews and descriptions of techniques used. We then look at what this research will lead to in the future and what we can expect in years to come. Finally, we conclude this article by reviewing what has been discussed.

BACKGROUND

More and more digital audio-visual content is now available online. Also more access networks are available for the same network different devices (with different resources) that are being introduced in the marketplace. Structured multimedia content (even if that structure is still limited) increasingly needs to be accessed from a diverse set of networks and terminals. The latter range (with increasing diversity) from gigabit Ethernet-connected workstations and Internet-enabled TV sets to mobile video-enabled terminals (Figure 1) (Pereira & Burnett, 2003).

Adaptation is becoming an increasingly important tool for resource and media management in distributed multimedia systems. Best-effort scheduling and worst-case reservation of resources are two extreme cases, neither of them well-suited to cope with large-scale, dynamic multimedia systems. The middle course can be met by a system that dynamically adapts its data, resource requirements, and processing components to achieve user satisfaction. Nevertheless, there is no agreement about questions concerning where, when, what and who should adapt (Bormans et al., 2003).

On deploying an adaptation technique, a lot of considerations have to be done with respect to how to realise the mechanism. Principally, it is always useful to make the technique as simple as possible, i.e., not to change too many layers in the application hierarchy. Changes of the system layer or the network layer are usually always quite problematic because deployment is rather difficult. Generally, one cannot say that adaptation technique *X* is the best and *Y* is the worst, as it highly depends on the application area.

The variety of delivery mechanisms to those terminals is also growing and currently these include satellite, radio broadcasting, cable, mobile, and copper using xDSL. At the end of the distribution path are the users, with different devices, preferences, locations, environments, needs, and possibly disabilities.

In addition the processing of the content to provide the best user experience may be performed at one location or distributed over various locations. The candidate locations are: the content server(s), any processing server(s) in the network, and the consumption terminal(s). The choice of the processing location(s) may be determined by several factors: transmission bandwidth, storage and computational capacity, acceptable latency, acceptable costs, and privacy and rights issues (see Figure 2).

Present adaptation technologies concerning content adaptation mainly focus on the adaptation

Figure 1. Different terminals access multimedia content through different networks

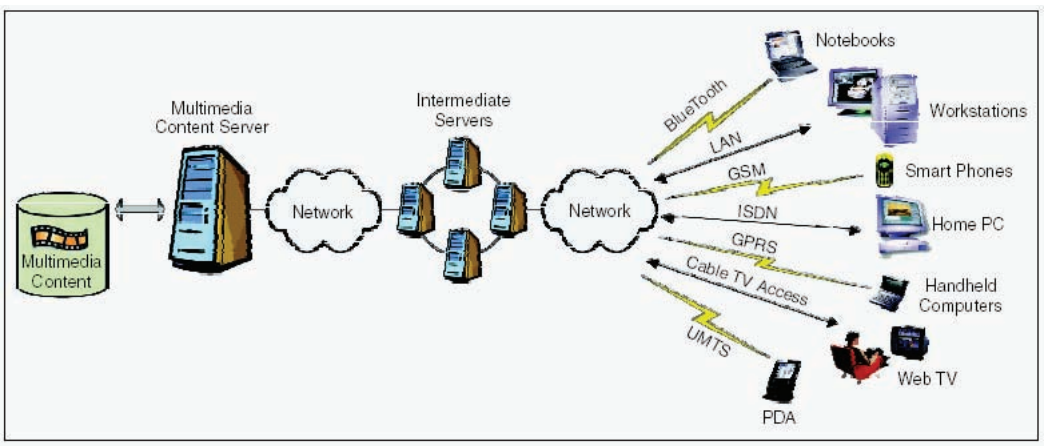
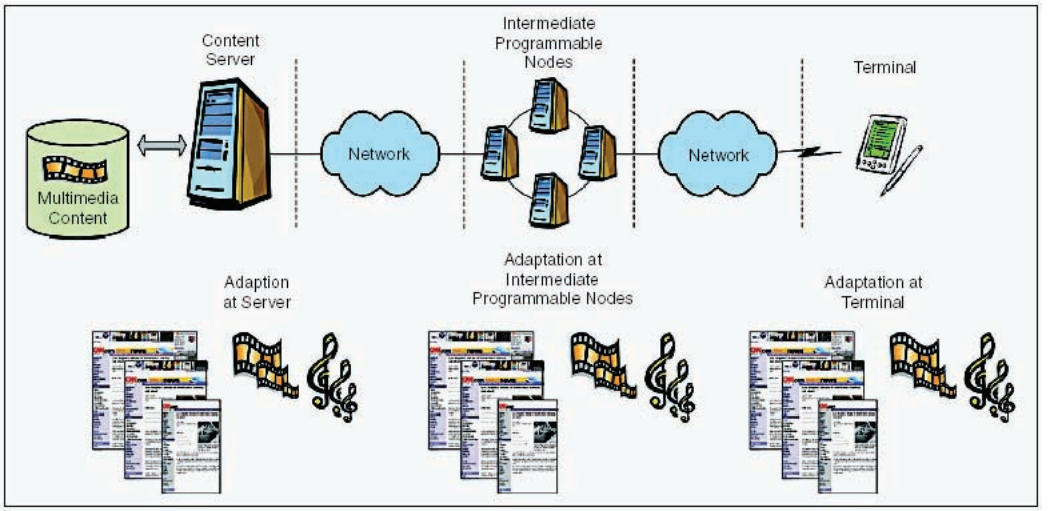


Figure 2. Adaptation may be performed at different places



of text documents. Therefore, one text document will be adapted on demand to the capabilities of different devices or applications. To fulfill this functionality the structure of the content must be separated from its presentation, i.e., the source document is structured using XML (Extensible Markup Language) and then dynamically processed to generate a presentation tailored to the available resources. One possible use case scenario will be to present the same information either on a standard Web browser or a WAP browser.

Efficient adaptation requires that the participating components know from each other and take advantage of adaptation steps done by other components, which needs standardised media, metadata, and communication. Several standardisation bodies (W3C, MPEG, and WAP) have already been established or are currently under development, which have recognised the need to create a framework that facilitates the efficient adaptation of content to the constraints and preferences of the receiving end.

MPEG-7 (ISO/IEC 15938-5:2002) provides tools for content description, whilst capability description and negotiation is provided for with CC/PP (Composite Capabilities/Preference Profiles, 2003) and UAProf (WAG User Agent Profile, 2001). MPEG-21 (ISO/IEC JTC 1/SC 29/WG 11), the "multimedia framework" includes Digital Item Adaptation (DIA), which enables standard communication of dynamic adaptation of both media resources and meta-data, enabling negotiation of device characteristics and QoS parameters. (Böszörményi et al., 2002)

In this section, the reasons for the need for interoperable and efficient multimedia content adaptation have been introduced. A number of standards groups (such as W3C, MPEG and WAP) that facilitate multimedia content adaptation by concentrating on the adaptation of associated XML-type documents have also been mentioned. The next section delves into the different technologies that help make up this exciting field.

MAIN THRUST OF THE CHAPTER

In this section we will look at the main themes found in multimedia content adaptation. We start with a look at a multimedia content adaptation architecture, a discussion on the present state of affairs regarding scalable coding and transcoding,

an analysis of the effect the actual location point the adaptation takes place and a brief summary of the relevance of user profiling.

Multimedia Content Adaptation Architecture

The networking access paradigm known as Universal Multimedia Access (UMA) refers to the way in which multimedia data can be accessed by a large number of users/clients to view any desired video stream anytime and from anywhere. In the UMA framework, multimedia information is accessed from the network depending on the following three parameters: channel characteristics, device capabilities, and user preference.

Figure 3 gives an example of different presentations (to suit different capabilities such as formats, devices, networks, and user interests) of the same information.

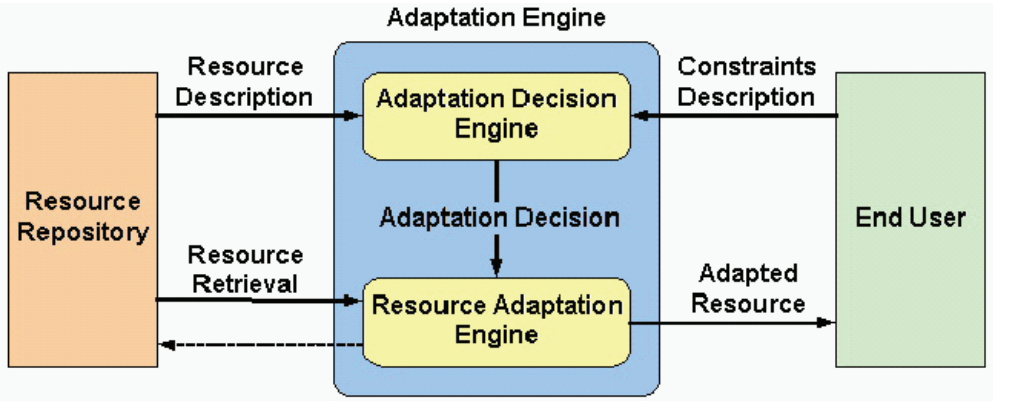
One option for UMA is to provide different variations of the content with different quality, bit rate, media modality (e.g., audio to text), etc. The problem with this option is that it is not too efficient from the viewpoint of variation generations and storage space. On the other hand, real-time transformation of any content implies some delay for the processing and a lot of computing resources at the server (or proxy server) side. Pursuing either of these two options assumes the use of an adaptation engine. Figure 4 gives a bird's eye-view of such an adaptation engine architecture that is applicable to the adaptation of any type of content.

The architecture consists of an adaptation engine that can be located on the server, an intermediate network device such as a gateway, router, or proxy, or even on the client. This engine comprises of two logical engine modules, the adaptation decision engine and the resource adaptation engine. The adaptation decision engine receives the metadata information about the available content (context, format, and adaptation options) from the resource repository and the constraints (terminal and network capabilities, user characteristics, and preferences) from the receiving side. If there are multiple versions of the content pre-stored in the repository and one of these versions matches the constraints, then this version is selected, retrieved, and sent to the end user. However, if the available resource does not match the constraints, but can be adapted, then the adaptation decision engine

Figure 3. Different presentations of the same information

Workstation/LAN	PC/Dialup	TV Browser	Gray PDA	BW PDA	Text Browser
					"bridge"
38 KB	23 KB	8 KB	4 KB	0.6 KB	0.01 KB
24 bit color	24 bit color	256 colors	4 bit gray	B/W	-
256 x 256	192 x 192	128 x 128	96 x 96	64 x 64	-
22 sec	13.5 sec	4.7 sec	2.4 sec	0.35 sec	0.01 sec

Figure 4. Bird's-eye view of an adaptation engine architecture



determines the optimal adaptation for the given constraints and passes this decision to the resource adaptation engine. The resource adaptation engine retrieves the resource from the repository, applies the selected adaptation, and sends the adapted resource to the end user.

Constraints can be grouped into four broad categories: user and natural environment characteristics, terminal capabilities, and network characteristics. The terminal and network constraints will set an upper bound on the resources that can be transmitted over the network and rendered by the terminal. Information like the network's maximum bandwidth, delay and jitter, or the terminal's resolution, buffer size, and processing power, will help the adaptation engine determine the optimal version of the resource for the given network and terminal device. As the user is the actual consumer (and judge) of the resource, user-related information, including user preferences, user demographics, usage history and natural environment, is equally important in deciding which resource should be delivered to the terminal.

The adaptation engine needs to have sufficient information about the context and the format of the multimedia resources in order to make a decision whether the resource is suitable for the user or how it should be adapted in order to offer the user the optimal context and quality. The description should therefore include information on the resource type, semantics, available adaptation options, and characteristics of the adapted versions of the resource (Panis et al., 2003).

Scalable Coding

If content is scalable, then adapting content may be done using scalable coding. This removes or alters parts of resources in such a way as to reduce their quality in order to satisfy the receiver's capabilities and needs. Currently available scaling options depend on the coding format to be used. The Holy Grail of scalable video coding is to encode the video once, and then by simply truncating certain layers or bits from the original stream, lower qualities, spatial resolutions, and/or temporal resolutions could be obtained (Vetro, 2003).

Current scalable coding schemes fall short of this goal. MPEG-4 is currently the content representation standard where the widest range of scalability mechanisms is available, notably in terms of data types, granularities, and scalability domains (Pereira & Ebrahimi, 2002).

Transcoding

This is another more complex option that typically refers to transforming the resource from one

coding format into another one, i.e., decoding the resource and encoding the resource using another codec (e.g., transcode an MPEG-2 video to an MPEG-1 video). According to Sun et al. (2003), the key design goals of transcoding are to maintain the video quality during the transcoding process and to keep complexity as low as possible. Cavallaro et al. (2003) identify three main approaches to video transcoding: content-blind transcoding, semantic transcoding, and description-based transcoding:

- **Content-blind transcoding** does not perform any semantic analysis of the content prior to conversion. The choice of the output format is determined by network and appliance constraints, independent of the video content (i.e., independent of the way humans perceive visual information). The three main content-blind transcoding categories are spatial conversion, temporal conversion, and colour-depth reduction.
- **Semantic (or intramedia) transcoding** analyses the video content prior to conversion. An example of such analysis is the separation of the video content into two classes of interest, namely foreground and background. Once this separation has been accomplished, the two classes can be coded differently to better accommodate the way humans perceive visual information, given the available network and device capabilities.
- **Description-based (or intermedia) transcoding** transforms the foreground objects extracted through semantic segmentation into quantitative descriptors. These quantitative descriptors are transmitted instead of the video content itself. In this specific case, video is transformed into descriptors so as to produce a textual output from the input video. Such textual output can be used not only for transcoding, but also for annotating the video content and for translating the visual content into speech for visually impaired users. This transformation is also referred to as cross-media adaptation.

Location of Adaptation

As well as the technologies used, the location used for multimedia content adaptation needs to be addressed. Resource delivery and adaptation can be sender-driven, receiver-driven, or network-driven.

Sender-driven proceeds to adapt resources at the sender/server node depending on the terminal and/or network capabilities received beforehand. After successful adaptation the sender transmits the adapted version of the resource to the receiver. This action requires a serious amount of computational power at the server node and goes at the expense of latency between the receiver's request and the server's delivery.

In contrast, the receiver-driven approach decides what and how to adapt at the terminal side although the real adaptation could take place somewhere else, e.g., on a proxy node. Adaptation directly at the end node could fail due to insufficient capabilities. Additionally network bandwidth will be wasted, too. Nevertheless, adaptation on terminal devices should not be strictly excluded.

The pure network-driven approach is transparent where the network, i.e., the transport system, is responsible for adaptation only. Typical use case scenarios will cover all kind of adaptation approaches described so far, i.e., resource adaptability along the delivery chain, from resource provider to resource consumer. A high-performance server node will provide some kind of preprocessing in order to facilitate easy adaptation along the delivery chain across a wide range of network and terminal devices. Network nodes such as routers or gateways will then perform so-called light-weight adaptations using segment dropping or minor editing techniques whereas proxy nodes could utilise more complex adaptation techniques.

Such complex adaptation techniques include not only scaling but also transcoding and cross-media. An adaptive terminal device could perform adjustments due to user and/or usage preferences. The complexity of these adaptations to be done in terminal devices depends on its capabilities, e.g., display resolution, computational power, local storage capacity, and buffer size.

User Profiling

In order for the personalisation and adaptation of multimedia content to take place, the users' preferences, interests, usage, and environment need to be described and modelled. This is a fundamental realisation for the design of any system that aims to aid the users while navigating through large volumes of audio-visual data. The expectation is that by making use of certain aspects of the user model, one can improve the efficacy of the system and further help the user (Kobsa et al., 2001).

This section described the components needed to make up a multimedia content adaptation architecture: transcoding, scalable coding, location of adaptation, and user profiling. The next section discusses the future of multimedia content adaptation by looking at UMA, transcoding, and scalability, specifically.

FUTURE TRENDS

The major problem for multimedia content adaptation is to fix the mismatch between the content formats, the conditions of transmission networks, and the capability of receiving terminals. A mechanism for adaptation needs to be created for this purpose.

Scalable coding and transcoding are both assisting in this. It can be seen that scalable coding and transcoding should not be viewed as opposing or competing technologies. Instead, they are technologies that meet different needs regarding multimedia content adaptation and it is likely that they will coexist.

Looking to the future of video transcoding, there are still quite a number of topics that require further study. One problem is finding an optimal transcoding strategy. Given several transcoding operations that would satisfy given constraints, a means for deciding the best one in a dynamic way has yet to be determined. Another topic is the transcoding of encrypted bit streams. The problems associated with the transcoding of encrypted bit streams include breaches in security by decrypting and re-encrypting within the network, as well as computational issues (Vetro, 2003).

The inherent problem with cross-media (description-based) adaptation is in preserving the intended semantics. What are required are not the blindfolded exchange of media elements and fragments, but their substitution by semantically equivalent alternatives. Unfortunately, current multimedia authoring tools provide little support for producing annotated multimedia presentations. Richly annotated multimedia content, created using document-oriented standards, such as MPEG-7 and MPEG-21 DIA, will help facilitate sophisticated cross-modal adaptation in the future.

For the implementation of UMA, "universal," scalable, video-coding techniques are essential components. Enhancements to existing video-coding schemes, such as MPEG-4 FGS (Fine-Granular-Scalability) and entirely new schemes will help drive the UMA ideal. More efficient FGS-encoders, tests on the visual impact of variability and more improved error resilient techniques are improvements that can be made to scalable coding schemes.

While some technologies such as content scalability and transcoding are fairly well established, there are still vital technologies missing for a complete multimedia content adaptation system vision. Many of these technologies are directly related to particular usage environments. While multimedia adaptation for improved experiences

is typically thought of in the context of more constrained environments (e.g., mobile terminals and networks), it is also possible that the content has to be adapted to more sophisticated environments, e.g., with three-dimensional (3-D) capabilities. Whether the adaptation processing is to be performed at the server, at the terminal, or partially at both, is something that may have to be determined case–by-case, depending on such criteria as computational power, bandwidth, interfacing conditions, and privacy issues.

CONCLUSION

The development and use of distributed multimedia applications is growing rapidly. The subsequent desire for multimedia content adaptation is leading to new demands on transcoding, scaling, and, more generally, adaptation technologies. Metadata-based standards, such as MPEG-7 and MPEG-21, which describe the semantics, structure, and the playback environment for multimedia content are breakthroughs in this area because they can assist more intelligent adaptation than has previously been possible.

A prerequisite for efficient adaptation of multimedia information is a careful analysis of the properties of different media types. Video, voice, images, and text require different adaptation algorithms. The complex nature of multimedia makes the adaptation difficult to design and implement. By mixing intelligence that combines the requirements and semantic (content) information with low-level processing, the dream of UMA could be closer than we envision.

REFERENCES

Bormans, J., Gelissen, J., & Perkis, A. (2003). MPEG-21: The 21st century multimedia framework. *IEEE Signal Processing Magazine*, 20(2), 53- 62.

Böszörményi, L., Doller, M., Hellwagner, H., Kosch, H., Libsie, M., & Schojer, P. (2002). Comprehensive Treatment of Adaptation in Distributed Multimedia Systems in the ADMITS Project. ACM International Multimedia Conference, 429-430.

Cavallaro, A., Steiger, O., & Ebrahimi, T. (2003). Semantic segmentation and description for video transcoding. Paper presented at the Proceedings of the 2003 International Conference on Multimedia and Expo, 2003, ICME '03..

Composite Capabilities/Preference Profiles. (2003). Retrieved from the World Wide Web March 2003 at: http://www.w3.org/Mobile/CCPP/

Extensible Markup Language (XML). (2003). 1.0 (3rd Edition). Retrieved from the World Wide Web October 2003 at: www.w3.org/TR/2003/PER-xml-20031030/

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. *CiteSeer.*

ISO/IEC. (2002). ISO/IEC 15938-5: 2002: Information Technology—Multimedia Content Description Interface—Part 5: Multimedia Description Schemes.

ISO/IEC (2003). ISO/IEC JTC 1/SC 29/WG 11: MPEG-21 Digital Item Adaptation Final Committee Draft. Document N5845, Trondheim, Norway. Retrieved from the World Wide Web July 2003 at: http://www.chiariglione.org/mpeg/working documents.htm#MPEG-21

Panis et al. (2003). Bitstream Syntax Description: A Tool for Multimedia Resource Adaptation within MPEG-21, EURASIP Signal Processing. *Image Communication*, Special Issue on Multimedia Adaptation, 18(8), 721-74

Pereira, F., & Burnett, I. (2003). Universal multimedia experiences for tomorrow. *Signal Processing Magazine*, IEEE, 20(2), 63-73.

Pereira, F., & Ebrahimi, T., (2002). *The MPEG-4 Book*. Englewood Cliffs, NJ: Prentice-Hall.

Sun, H., Vetro, A., & Asai, K. (2003). Resource Adaptation Based on MPEG-21 Usage Environment Descriptions. Proceedings of the IEEE International Conference on Circuits and Systems, 2, 536-539.

Van Beek, P., Smith, J. R., Ebrahimi, T., Suzuki, T., & Askelof, J. (2003). Metadata-driven multimedia access. *Signal Processing Magazine*, IEEE, 20(2), 40-52.

Vetro, A. (2003). Visual Content Processing and Representation, Lecture Notes in Computer Science, (pp. 2849). Heidelberg: Springer-Verlag.

WAG. (2001.) User Agent Profile. Retrieved October 2001 from the World Wide Web at: http://www1.wapforum.org/tech/documents/WAP-248-UAProf-20011020-a.pdf

KEY TERMS

Bit Stream: The actual data stream, which is the transmission of characters at a fixed rate of speed. No stop and start elements are used, and there are no pauses between bits of data in the stream.

Content Scalability: The removal or alteration of certain subsets of the total coded bit stream to satisfy the usage environment, whilst providing a useful representation of the original content.

Cross-Media Adaptation: Conversion of one multimedia format into another one, e.g., video to image or image to text.

Multimedia Content Adaptation: The process of adapting a multimedia resource to the usage environment. The following factors make up this usage environment: users preferences, device, network, natural environment, session mobility, adaptation QoS, and resource adaptability.

Transcoding: The process of changing one multimedia object format into another.

UMA: How users can access the same media resources with different terminal equipment and preferences.

User Modelling: In the context of adaptation, the describing/modelling of the users preferences, interests, usage, and environment.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 4, edited by M. Khosrow-Pour, pp. 2051-2057, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.10
Introduction to Mobile Multimedia Communications

Gour C. Karmakar
Monash University, Australia

Laurence S. Dooley
Monash University, Australia

Michael Mathew
Monash University, Australia

ABSTRACT

In order to meet the ever increasing demand by people using mobile technology and its associated services based on multimedia elements in addition to voice, mobile communication technologies has since evolved from analog to digital and 1G to 4G. This chapter presents a contemporary review of all generations of mobile communication technologies, including their standards. 1G to 3G mobile communication technologies are mainly optimised for voice communication, using circuit switched networks. To provide high transmission mobile services at low cost in all levels of networks—personal, home, and cellular—it is imminent to exploit the merits of all existing technologies such as Bluetooth, WLAN, and HiperLAN, and use IP as a backbone network in 4G mobile communication standards. The key research challenges for mobile terminals, systems, and services for 4G networks are also presented in this chapter.

INTRODUCTION

Multimedia refers to the combination of different types of media elements such as text, audio, image, and video in a digital form which is represented and manipulated by a single electronic device or a single computing platform such as a PC (Chapman & Chapman, 2000). Interactive multimedia provides the interaction facility with users so they can access and exit the system as they wish.

Text comprises a string of alphabets from a particular character set. Image is a visual object

consisting of a rectangular pattern of dots or primitive elements—lines, curves, circles, and so on (Halsall, 2001). Examples of images are computer generated graphics and digitized documents, pictures, and graphic arts. Images could be either 2D or 3D. Both text and images are inherently in a digital form. Bandwidth requirement for the transmission of text and image is less than that of high-fidelity audio and video. Audio is generally represented by the amplitude of sound waves, which includes low-fidelity speech during telephonic conversation, high-fidelity CD-quality audio and surround sound.

Video is a sequence of still images or frames displayed in a repaid succession so the human eye cannot pick up their transitions, and hence it creates an illusion of motion. Examples of video include movies, short films, and animation. Both audio and video are continuous time varying signals and analog in nature. For an integrated representation, all media elements must be represented in digital form.

With the rapid growth of the Internet, multimodal representation and interactive facilities of multimedia-based systems, the necessity of required information, accurate presentation, and quick perception, the applications of multimedia are burgeoning in every aspect of human life ranging from home, education, medicine, e-commerce and m-commerce, to airport security. The demands of these applications are met with the different types of multimedia-based services that are a combination of a number of media elements, but not limited to—interactive television (text, audio, and video), video phone and conference (speech and video), computer supported cooperative working (text and images), on-line education (text, images, audio and video), multimedia electronic mail (text, images, and audio, for example), e-commerce (text, images, audio, and video) and Web and Mobile TV (text, audio, and video) (Halsall, 2001).

Due to the inherent high data rates, especially for audio and video, a number of compression techniques for both audio and video have been introduced, which has made it possible to transmit the video over broadband networks. Even with the advancement of compression technologies, the usual bit rate for a speech signal is 64kbps, while it is 384Kbps for low quality video, and up to 2Mbps for high quality video (Sawada, Tani, Miki, & Maruyama, 1998). The tentative bandwidth requirement for the future networks for enhanced-reality multimedia communications are projected from 1Mbps to 30Gbps for 3D audio and 3D video (Ohya & Miki, 2005).

Requirements for an on-demand real-time telephone network access for various purposes such as business, education, and social, cultural, and psychological factors, the scope of technology has been expanded from land-base fixed communications, to wireless and mobile multimedia communications. With the advent of wireless and mobile networks, people now have the opportunity to communicate with anybody, anywhere, and at any time. The number of mobile users is rapidly increasing all over the world (Rao & Mendoza, 2005; Salzman, Palen, & Harper, 2001). It is estimated the number of worldwide mobile users will be 1.87 billion by the end of 2007 (Garfield, 2004).

To meet the ever-increasing consumer demands and make the multimedia-based services as appealing as possible, the mobile networks have evolved from 1G to 4G. 1G mobile networks were commercially released in 1980. Examples of a 1G network include Advanced Mobile Phone Service (AMPS) and Total Access Communication System (TACS). An AMP is in American roaming, while TACS is European roaming. These are based on cellular analog technology. They were initially provided voice service at a rate of 2.4 kbps (Casal, Schoute, & Prasad, 1999), which had been extended to 19 kbps with the introduction of Cellular Digital Packet Data (CDPD) in the bedrock of 1G analog cellular networks for providing digital data service. Since all of these 1G networks are analog broadcasting, they are

not secured and hence vulnerable to intercepting calls through a radio frequency scanner (McCullough, 2004).

For providing secured communication, higher transmission of up to 64 kbps, better quality of signal, as well as meeting ever increasing demand from users, 1G analog mobile standards evolved into 2G digital communication technologies around 1990. 2G mobile communication standards include Global System for Mobile Communication (GSM), Digital AMPS (D-AMPS), cdmaOne (IS-95), Personal Digital Cellular (PDC). Later, some of 2G standards have been enhanced to provide higher bit rate transmission up to 384 kbps, and data services including the Internet, which is known as 2.5G. An example of such an enhancement of GSM is a circuit switched network called GPRS, which is a packet switched network, and provides Internet facilities for mobile users (Andersson, 2001; Wikipedia, 2006).

To increase the capacity up to broadband communication, that is, 2 Mbps, and hence the number of subscribers, the specification of the 3G standard is outlined by IMT2000 considering the hierarchical cell structure, global roaming, and dynamic allocation of a radio frequency from a pool. This enables 3G to provide services such as fast Internet browsing and mobile multimedia communications—mobile TV, video phone, and video conferencing. Despite 3G being adopted as a current standard, its full-fledged implementation is being delayed because of the problems for the introduction of new hardware and dynamic frequency allocation—consequently making 3G obsolete and introducing 4G.

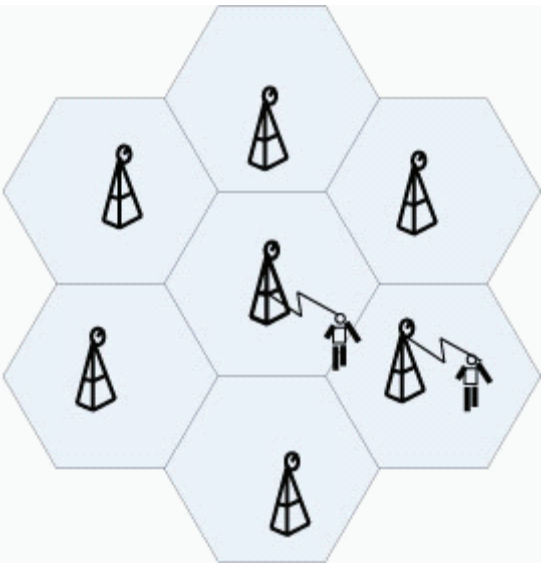
It is expected 4G mobile communication networks will be commercially available in the market within 2010. Its main purpose is to provide high transmission of multimedia data up to 1Gbps based on IP-core networks (Hui & Yeung, 2003; Ohya & Miki, 2005). It will consist of heterogeneous state-of-the-art technologies ranging from ad-hoc sensor networks, to intelligent home appliances, and to provide cutting-edge services including ubiquitous, reality, and personalised communications.

GENERATIONS OF MOBILE MULTIMEDIA NETWORKS

The first mobile telephone standard was introduced by AT&T in St. Louis, Missouri, USA, and was available for public users of this city in 1946. The mobile system (a radio telephone network) was installed on the top of a building in the metropolitan area. All users received their services through this system with a mobile operator, who routed both incoming and outgoing calls to the intended recipients. A high powered radio-based transmitter was used to cover the required area. However, the number of frequencies and hence the channels were very limited. It was difficult to have a dial tone and users had problems to complete calls. The system operated in a half-duplex communication mode, that is, one way communication at a time based on push-to-talk communication mode. The mobile set was very heavy and bulky.

An example of one of the initial mobile telephone standards is the Mobile Telephone Systems (MMT). Because of the demand beyond the capacity and previously mentioned disadvantages, the traditional analog mobile telephone technology was upgraded in 1960, and an Improved Mobile Telephone Service (IMTS) was introduced, which eliminates the push-to-talk mode by providing direct access to the dial telephone networks and full-duplex transmission. In this standard, the usage of high powered radio at 200 watts created interference up to 100 miles, which allowed limited available channel capacity and frequency reuse. The limited channel capacity and excessive demands for communication at anytime from anywhere brought the birth of cellular mobile telephone networks in 1974 (Bates, 1995). In cellular technology, a coverage area is divided into small hexagonal zones known as cells shown in Figure 1.

Figure 1. Organization of hexagonal cells for a coverage area



Each cell has its base station consisting of a transmitter and two receivers per channel, a base station controller, an antenna system, and a data link to the base station controller or cellular office. The average size of a cell is three to five miles, which eventually resulted in a far less coverage area compared with non-cellular technology for a mobile phone, and hence reduced the required transmission power and interference between the neighbouring cells significantly. This made it possible to reuse the frequency and to develop a protection zone in the form of clustering of frequencies for a particular cell against the interference of its neighbouring cells, which paved the way of increasing the channel capacity of a cell, and hence the mobile communication networks. In order to meet the ever increasing demand for communication at anytime for anywhere using any technology, the cellular mobile communication technology has evolved so far into four generations shown in Table 1.

The description of each generation, including their related network standards are described in the following section.

First Generation (1G) Cellular Mobile Networks

Because of the increasing demands of users, it was necessary to increase the capacity and mobility of mobile telephone users. To meet this demand, the analog cellular technologies were designed and completed by AT&T in 1974 and commercialised in the 1980s. This analog service is still available around the world and played a vital role to increase the popularity of mobile phone and the mobile users at 30 to 50% per year, which contributed to approximately 20 million users within 1990 (IFC, 2005). Examples of 1G generation mobile cellular networks are Advanced Mobile Phone Service (AMPS), Total Access Communication System (TACS), and Nordic Mobile Telephone (NMT).

Advanced Mobile Phone Service (AMPS)

AMPS is the first analog cellular system, which was designed and developed by AT&T in the

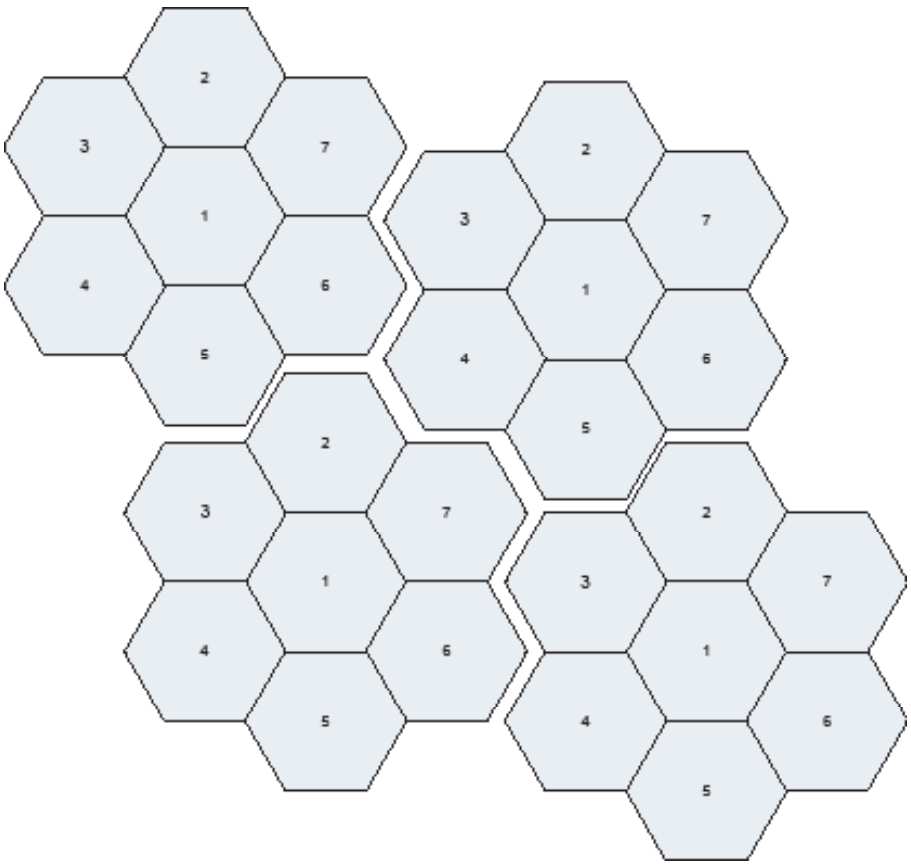
Table 1. Generations of cellular mobile communication networks

Year	Network Standards	Services	Capacity
1980's	1G Analog AMPS TACS NMT	Voice Small data	2.4 kbps – 9.6 kbps
1990's	2G Digital GSM D-AMPS PDC cdmaOne CSD iDEN 2.5 G or 2G+ Digital GPRS HSCSD WiDEN EDGE CMDA2000 1xRTT	Voice SMS Some data Voice Data Email Internet	Up to 48.6 kbps Up to 384 kbps
2000's	3G Digital W-CDMA CDMA2000 1xEV TD-SCDMA	Fast Internet Mobile multimedia Video Phone Mobile TV Remote access	Up to 2Mbps
2010's	4G Digital Wi-Fi UMB UMTS Revision 8 WiMAX WiBro	Mobile Ubiquitous Reality communications (e.g., virtual reality and 3D applications) Multiple services at the same time Personalised communications	Up to 1Gbps

1970s in the Bell Labs as a research project, and commercially launched in America in 1984. It was one of the revolutionary achievements in the mobile communication technologies due to the smaller area coverage and low powered output through the use of the cellular technology. The low powered output produces lower interference, and hence it effectively allows frequency reuse in the form of cell clusters. The cells in a cluster may interfere with their neighbouring cells. To produce minimum interference, frequency reuse must be at least two cells apart. AMPS generally uses either 12 or 7-group cell patterns (Wesel, 1998). An example of a 7-group cell pattern with frequency repetition and separation among them is shown in Figure 2.

This directly yielded a higher capacity system compared with a traditional non-cellular analog mobile system. The division of area coverage into hexagonal cellular regions also provided the

Figure 2. 7-group cell pattern with frequency repetition and separation



mobile users with a greater mobility. AMPS uses frequency division multiple accesses (FDMA) for voice channels and operates in 800-900 MHz. Initially in 1984, the Federal Communications Commission (FCC) allocated 666 channels (333 channels for each carrier) for two carriers—A and B. Lower frequencies were given to the license B, which was owned by wireline companies, while the upper frequencies were allotted to the license A, owned by non-wireline companies. Due to the rapid growth of the mobile cellular users, in 1989 the FCC had increased the channels from 333 to 416 for each carrier, that is, 832 in total (Wikipedia, 2006). Twenty-one of 416 channels are control channels that carry control data, and the remaining 395 channels are used for two way voice communications.

Despite its analog nature, AMPS has countrywide coverage, it is not possible to encrypt and decrypt messages, and therefore, anyone can intercept the call through a handheld radio frequency scanner. Hence, it is not secure at all. Other disadvantages are: the lifetime of the battery is not long enough, poor quality of the transmitted message, and call charge is costly (ePanorama, 2006). For these reasons, AMPS is being replaced with 2G standards such as Digital AMPS (D-APMS), GSM, and so on.

Total Access Communication System (TACS)

Total Access Communication System (TACS) was introduced in Europe in the late 1970s, and

modified after AMPS using FDMA. Therefore, it is an analog mobile communication standard and is known as the European version of AMPS. It is used in the UK, China, and Japan (ePanorama, 2006). Extended TACS called ETACS was introduced in the UK by extending the channel capacity of TACS. Both TACS and ETACS are completely outdated in Europe and replaced by GSM (Wikipedia, 2006).

Nordic Mobile Telephone (NMT)

As with AMPS and TACS, Nordic Mobile Telephone (NMT) is another analog mobile telephone standard and was widely used in Nordic countries. It was introduced in the late 1970s in collaboration with Finland, Sweden, and Norway. There were two versions of NMT—NMT-450 and NMT-900. NMT-450 used 450 MHz frequency, while NMT-900 used 900 MHz. NMT is also completely outdated, and hence it has been replaced by GSM and W-CDMA (Wikipedia, 2006).

Second Generation (2G) Cellular Mobile Networks

The limitations of analog cellular networks have been alluded to in the previous section. To remove these limitations, meeting the ever increasing demand from users, and providing better quality, the countries and individual companies felt the importance for the introduction of digital cellular communication standards called 2G cellular mobile communication standards.

For digital encoding, the 2G standards are able to provide small data transmission in addition to voice communication. Some 2G standards can also provide a few extra advanced services: short messaging service (SMS) and enhanced messaging service (EMS). The digitisation of voice makes it possible to apply filtering or audio compression techniques to reduce the size of the transmitted voice, which leaves some room to multiplex more subscribers into the same radio frequency channel, and produce better quality compared with analog cellular standards by reducing noise. Examples of 2G cellular mobile communication standards include Global System for Mobile Communication (GSM), Digital AMPS (D-AMPS), cdmaOne (IS-95), and Personal Digital Cellular (PDC).

Global System for Mobile Communication (GSM)

Initially, to address many compatibility problems among many digital radio technologies of the European countries, the GSM-Group Special Mobile was formed by the Conference of European Posts and Telegraphs (CEPT) in 1982. After this, the full name of GSM was changed to Global System Mobile Communications, and its responsibility was handed over to the European Telecommunication Standards Institute (ETSI) in 1989 (ePanorama, 2006). The technical specification of GSM was completed in 1990, and it was commercially launched in 1991 from Finland.

From its initial journey, it evolved itself as the most popular standard, and took over 70% of the mobile market with 1.6 billions subscribers in 210 countries all over the world by 2005 (Wikipedia, 2006). The ever increasing popularity of GSM was due to the following main reasons:

- GSM is based on an open system architecture. Therefore, any operator from any country can provide its services using their available hardware.
- It was made compulsory by rule to use GSM in European countries to increase the interoperability.
- GSM first introduced SMS text messaging and pre-paid accounts which are very popular among students and teenagers. Pre-paid accounts also dominated the market of developing countries, since it does not require any accountability or personal verification.

- It can provide a good quality service with moderate level security and can work under a wide range of frequencies.
- The cost is low.
- It can provide international roaming, that is, if a user is overseas, which is not covered by his/her own operator, the user can still receive service by the operator of the visiting location if there is an agreement between those two operators.

The majority of the GSM networks use 900 MHz or 1800 MHz frequency bands. There are some networks in USA and Canada that use 950 MHz or 1900 MHz frequency bands. In addition to these, GSM can also operate on 400 MHz and 450 MHz frequency bands. For multiplexing, it uses a combination of TDMA and FDMA, which allows the operator to allocate up to eight users, one for each time slot, into a single channel. The modulation technique used in GSM is Gaussian Minimum Shift Keying (GMSK), which effectively reduces the inter-channel interferences.

Like Integrated Services Digital Networks (ISDN), GSM can provide the data services—file or data transfer, and send a fax in up to a maximum speed of 9.6 kbps. GSM has been extended in order to increase its speed for providing faster data services. For example, High Speed Circuit Switched Data (HSCSD) is an expanded form of GSM, which can transmit up to 57.6 kbps by using multiple (e.g., eight) time slots for a single user or connection.

Digital AMPS (D-AMPS)

The digital form of the older analog AMPS standard is referred to as Digital AMPS (D-AMPS) or TDMA/IS-136. It is mainly used in the U.S. For transmitting information, it uses 824-849 MHz frequency band, while for receiving information, a frequency band of 869-894 MHz is utilised. The capacity of each radio frequency channel is 30 KHz. Each channel is divided into a number of time slots for squeezing several messages into a channel using the TDMA technique.

cdmaOne (IS-95)

cdmaOne is also known as IS-95. It is one of the 2G mobile communication standards and was originally introduced based on Code-Division Multiple Access (CDMA) multiplexing technology in 1993 by Qualcomm. The two versions of IS-95 are IS-95A and IS-95B (Mobile, 2007). The former operates in a single channel at 14.4Kbps, while the latter can support up to eight channels and 115 kbps.

Personal Digital Cellular (PDC)

Personal Digital Cellular (PDC) is a 2G Japanese digital cellular standard. After its introduction, it was widely used in Japan. This standard enforced Japanese to look into another alternative standard, as their old standard was incompatible with all other existing standards. For this reason, cdmaOne was gaining popularity in Japan in the late 1990s (Andersson, 2001).

Evolutions of GSM

As GSM is a circuit-switched mobile telephone network, it provides the limited amount of data services. In circuit-switched networks, dedicated connection is established for the entire life of a call and users usually need to pay per minute, whether they do or do not use the allocated connection. Another problem is users are charged as per usage time, and if it requires transmitting data through GSM, it needs to set up a dial-up connection. These issues acted as the driving force for the evolution of GSM, and consequently, it evolved itself into a packet switched network, namely General Packet Radio Services (GPRS) for providing the facilities of a data network Internet. An extra network, namely GPRS core network is built on the bedrock of the GSM network, collectively called GPRS

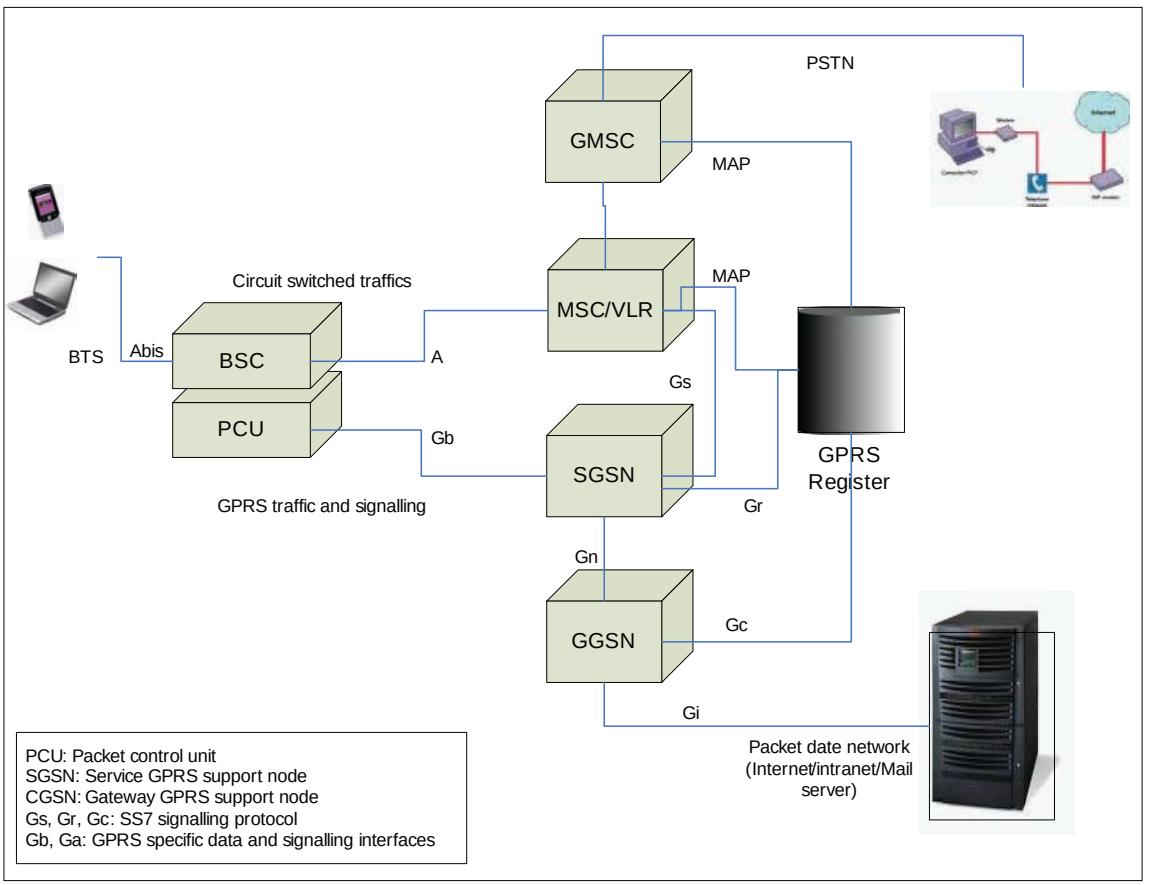
network. It allows GSM users to receive both vice and extra data services—mainly Internet browsing, e-mail, file and data transfer, and so forth. The architecture of the GPRS network is shown in Figure 3.

GPRS is regarded as 2.5 generation mobile network. GPRS is always online, that is, it provides the immediate data services whenever users require, just like a leased line or broadband Internet connection in which charges are based on data being transferred, or a flat rate provided by a particular operator. In contrast as mentioned previously, GSM usually charges based on the duration of a call, regardless of whether data is being transferred or not. The way of accessing both voice and data depends on the types of Capabilities—Class A, Class B, and Class C.

Class A provides the simultaneous access of both voice and packet data services. Class B provides voice and packet data services one at a time with a higher priority given to voice. If a voice call arrives, packet data services will be suspended, and resume back automatically at the end of the voice. Class B is the most common and is provided by most of the operators. Class C requires manual switching between voice and packet data by an operator (Wikipedia, 2006; Andersson, 2001).

There are eight TDMA time slots. The maximum capacity of each time slot is 20kbps. Therefore, theoretically GRPS can provide up to 160 kbps, however, this is limited by the multi slot classes (number of time slots used for downlinks, uplinks, and active slots), and the type of

Figure 3. The architecture of the GPRS network



encryption techniques used. The four encryption techniques used in GPRS are CS-1, CS-2, CS-3, and CS-4. CS-1 is the robust encryption technique, and hence it can accommodate 8kbps and cover 98% of the normal coverage, while CS-4 is the least robust encryption technique, which provides the data rate 20kbps and covers only 25% of the normal coverage area (Wikipedia, 2006).

Third Generation (3G) Cellular Mobile Networks

3G is a modern day mobile platform that has replaced the existing 2G standards. Some of the features of 3G articulated by IMT2000 include hierarchical cell structure, global roaming, and an expanding radio spectrum (Akyildiz, McNair, Ho, Uzunalioglu, & Wang, 1999). The claimed benefits of upgrading from the legacy networks mainly used for voice and SMS to 3G are endless. TV direct to your mobile, video calling, fast mobile Web browsing, and remote access are just the icing on the cake with the enormous potential of 3G. The most influential application for 3G was predicted to be video calling. With the deployment of 3G comes increased bandwidth, which opens a whole new world of possibilities.

The reasons for the delay of the rollout of the 3G network in some countries were due to:

- The costs involved in building new networks;
- Costs of additional spectrum licensing;
- Frequency distribution variation between 3G and 2G networks; and
- Costs of upgrading equipment.

However, Japan has been the leader for 3G on a commercial scale, with an expected extinction of the 2G network by 2006 (Hui & Yeung, 2003).

One of the 3G standards is Wideband Code Division Multiple Access (WCDMA) technology, a higher speed transmission protocol used in an advanced 3G system, designed as a replacement for the aging 2G GSM networks deployed worldwide. WCDMA provides a different balance of cost in relation to capacity and performance. WCDMA promises to achieve reduced costs and provide more value for day-to-day applications (Wikipedia, 2006).

3G broadband mobile communications makes access to sophisticated workplace technology inside the 3G handset even faster, making working life more flexible and developing still further the "virtual office" complete with e-mails, video conferencing, and high speed access to services without the daily commute. WCDMA works on a higher frequency in comparison to GSM and therefore travel a shorter distance.

Obviously, the real applications for 3G are the real-time ones. They include video telephony (video conferencing), video streaming, remote wireless surveillance, multimedia real-time gaming, video on demand, and more. These applications will drive usage and increase service-provider revenue. Consequently, they also will raise equipment sales (Hackett, 2001; UMTS World, 2003).

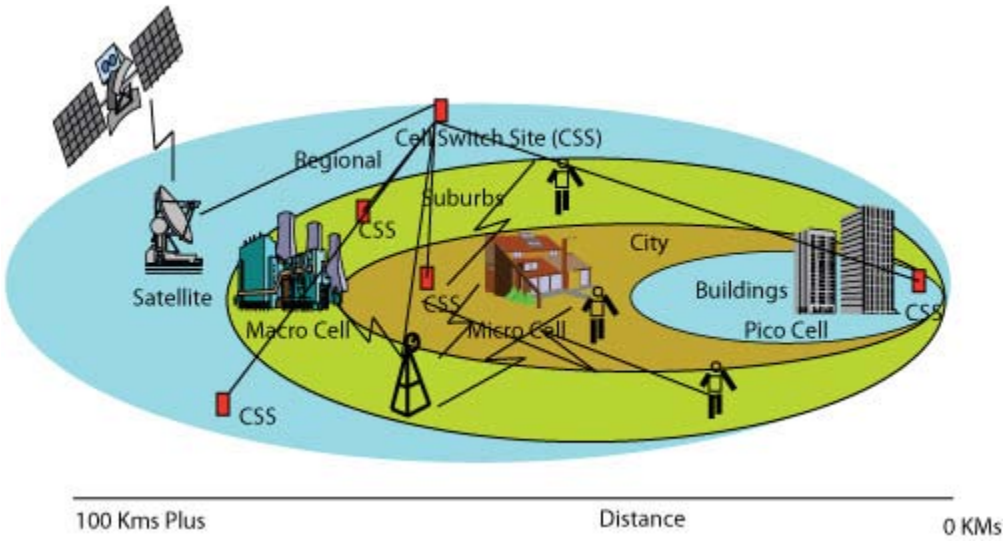
Hierarchical Cell Structure (HCS)

This portrays the different environments and cell structures across a given area.

The capacity is conversely proportional to coverage. The hierarchical cell structure shown in Figure 4 will cover all areas of the mobile user, thereby giving the user wider coverage and mobility starting from the high capacity pico-cell to the large global cell. Since the satellite takes part into the congestion control and it is a global extension to micro and macro cells, the capacity of 3G networks will increase, and hence it is capable of providing the services with more users (Akyildiz et al., 1999).

Figure 4 highlights the different operating environments for the mobile user. The mobile user will access wireless networks using the mobile terminal. This terminal uses radio channels to

Figure 4. The hierarchical cell structure



communicate with the base stations. This gives the user access to the terrestrial networks. As the distance increases the bandwidth decreases, with the bandwidth at its maximum in the pico-cell. However, although the quality of service (QoS) drops with distance, the coverage is transformed from base stations to satellite using the fixed Earth stations in areas of remote access. Finally, there are a number of cell site switch (CSS) that govern a number of base stations. These CSSs manage the radio recourses and provide mobility management control, location updates, and manage global roaming.

Global Roaming

The third generation wireless networks begin to implement terminal and personal mobility as well as service provider portability. Terminal mobility is the ability of a network to forward a call to a mobile terminal irrespective of its attachment point to the network. This means if the user is in an area of poor coverage the mobile will automatically shift to a service provider with strong coverage in the area giving the user good QoS. Personal mobility provides the users with an access to their personal services regardless of their attachment points to the network or mobile terminals (Akyildiz et al., 1999).

This level of global mobility freedom will be dependent on the coordination of a wide range of service providers to build a compatible backbone network and have strong agreements and cost structure. The first step is the development of global roaming agreements between different countries.

Radio Spectrum

Previously frequencies were allocated to specific sectors such as paging, cellular, mobile data, and private mobile radio. In 3G the radio spectrum is designed to standardize a pool of frequencies which could be managed to meet the global market needs and the technological developments. IMT2000 is designed to provide a spectrum of frequencies across different nations to meet with the demands of the mobile world, and hence it provides connections to heterogeneous backbone networks both wired and wireless.

Radio waves used to deliver 3G services are transmitted at a higher frequency than for 2G and travel a shorter distance. This may lead to coverage area or cell size for a 3G base station being smaller

than 2G site. Also, with the increase in demand from users in a particular cell, the size of that cell shrinks; the only way to ensure QoS is to overlap between cells. Therefore, the QoS in a high usage environment will decrease due to factors such as conjunction, interference, and static.

The location of cell sites is critical with 3G networks to avoid interference between adjacent cells, which in turn is one of the major issues to be addressed. Researchers are always looking to develop ways to better share the frequency and increase the bandwidth (Ariyoshi, Shima, Han, Karlsson, & Urabe, 2002). One of the solutions implemented is time division duplex (TDD). This application is designed to separate outward and return signals on the frequencies being used. TDD has a strong advantage in the case where the asymmetry of uplink and downlink data speed is variable (Wikipedia, 2006).

Another technique that is used is frequency division duplex (FDD), in which the uplink and downlink sub-bands are said to be separated by the frequency offset. FDD is more efficient in the case of symmetric traffic. FDD also makes radio planning easier and more efficient, as the base stations transmit and receive on different sub-bands. FDD is currently being used by the 3G network and works consistently, giving the user two separate frequencies to avoid interference (Wikipedia, 2006).

Key Characteristics and Application

This section will highlight some of the key applications offered as a result of the rollout of 3G. Each application will be described in relation to practicality in the industry sectors, and critical feedback on the deployment and operations will be discussed. In areas where problems have been identified, a brief overview will cover the cause of the problem with recommended solutions (UMTS World, 2003).

Fixed-Mobile Convergence (FMC)

Fixed-mobile convergence (FMC) is based on the use of dual-mode handsets. However, it is a broad area. For the purposes of this section, FMC means the converging functionality between desk phones and mobiles—set the mobile as a dual identity where the mobile can be addressed as a desk phone and mobile at the same time.

FMC essentially is a system designed to ensure corporate users can, for example, expect the same functionality from their PBX as their mobile and vice versa—same address book, call redirection, call transfer, message bank, and so on.

Some key characteristics of FMC include:

- Convergence of desk phone with mobile;
- Mobile convergence integrated with VoIP over the PABX;
- Functional transparency (address book, message bank, caller redirection, etc.);
- Anytime, any place access to employees; and
- Cost savings by organisation and clients.

The implementation of FMC has raised a few concerns in regards to employee security, PABX security, routing costs, and implementation of hardware that will enable a mobile to automatically convert to a part of the PABX and vice versa. The implementation of FMC is dependant on a few basic system characteristics such as:

1. Can a mobile have a dual identity?
2. What area of coverage will mobiles act as a part of the PABX and the security measure required?
3. What are the costs associated with implementation of hardware and software that will work over the two different systems?
4. Is it possible for the mobile to have the intelligence to recognize a call between the PABX and the mobile network?

The main advantage of FMC is the increased reach capacity of staff within the organisation. The other benefit is the cost saving structure for the firm routing calls via the PABX system.

Sales Force Automation

The sales force is the backbone of most organisations. 3G services are designed to introduce a new level of flexibility and convenience in the day to day life of the sales force. Figure 5 is an example of areas where 3G in conjunction with sales force automation (SFA) software can add value to the operations within organisation.

Some applications introduced to the sales force are:

1. Direct link from the mobile device (lap top or PDA to the company server);
2. Access to e-mail on the road;
3. Video calls; and
4. Wireless broadband service anywhere and anytime.

As 3G speeds increase in the future, and HS-DPA rollouts take place, the true potential of such mobile applications is likely to grow. The issue of bandwidth will be history and the connection problem will be rectified with permanently connected devices (Gartner, 2005).

SFA software will provide an interface that will link the user to different departments within the firm accessing real-time updates on order, inventory management, virtual conferences, and client data.

At the moment return on investment may be hard to forecast in terms of dollar value but as the systems develop the question will not be "how much does it cost to implement?" but rather "what are we missing by not have the application?" Some of the benefits driving the deployments are:

- Speed of business: agents can have access to client data immediately;
- Make changes in real time;
- Improving accuracy for logistic and inventory management;

Figure 5. Schematic diagram for mobile sales force automation

- Real-time solutions and access to real-time information; and
- Quick response time in coordinating within the organisation.

The current problem is the lack of custom SFA software in the marketplace. In cases where custom built SFA application are required, the cost seems to outweigh the benefits. There is also the case of overloading to many applications onto one small devices for it to be useful, which may turn it into a threat rather than a tool. The trick is getting companies to figure out what the key activities are and where the sales force will benefit from having mobile device access and focus on those.applications.

VoIP Mobile

VOIP integrated mobile services are still on the agenda. The use of the Internet as the sole platform to make calls is still not developed. The biggest benefit of mobile VoIP is having the ability to call over the Internet—in other words, creating huge cost savings. Unfortunately, this assumption is not true; the cost for organisations to set up mobile VoIP still seems to be expensive as it is in its infancy stages. That is not to say that the cost will gradually subside with developments in the field and time.

A number of companies are on the bandwagon trying to produce new developments to ensure quality of service. The two key factors in implementing successful VoiP is: (1) to increase the bandwidth; and (2) to ensure seamless flow of information from one point to another.

One in every five companies are looking into FMC—for most, it is cost savings rather than mobile VoIP, FMC is predicted to be the next big boom. It is believed that mobile VoIP will be an application running on 4G networks rather than the current 3G platforms.

Mobile TV

Mobile TV poses no direct benefit for the firms—companies are concerned it will be a distraction to the employees rather than a benefit. Mobile TV is targeted towards the mass audience rather than companies. For example, a certain target group would use this option to watch a football match or other similar programs.

The problem, however, is the end result—a fan or an enthusiast of a particular sport would prefer to watch the event on a big screen and with the ease of access to a number of venues. Mobile TV does not seem feasible. Businesses could incorporate mobile TV as a sales channel to run their advertising. For example, the GPRS on the phone might pick up a movie theater in the vicinity and automatically play the trailer of a movie.

Corporations will not spend money on unnecessary functionality such as mobile TV. To the general public, mobile TV is not a practical application due to a few reasons: (1) the end result (looking at a show on such a small screen); (2) the cost of usage; and (3) ease of access to substitute options to view programs.

Video Calling

According to the Shosteck Group, the size of the market for video messaging and calling will reach an estimated US$10 to US$28 billion by 2010, while in 2005, the Yankee Group estimated that already in 2007, video services will generate revenue of more than US$3.5 billion. Video calling was set to be the key element that would differentiate 3G from 2G. The modern user can see the direct benefits of video images and real-time interaction on their phones. There is now no doubt services based on video are emerging as a major source of operator revenue.

Nokia's own estimate is that the market for mobile entertainment and media services, including music, ring tones, mobile TV, and browsing, will have a value of around 67 billion euros in 2010.

This market will be flooded with a variety of video entertainment services (Le Maistre, 2002).

Until today, most users believe that 3G refers to video. This is because most service providers use the video calling as a flagship service in 3G launch, the person-to-person video call has already emerged as a key differentiator for 3G technology. Nokia studies show that as far as building 3G awareness among users go, the strategy has worked well in many markets. For example, according to research by Nokia in Hong Kong, Italy, Japan, and Korea in 2005, the vast majority of users in live networks named video calling as the key reason for subscribing to 3G (Orr, 2004).

However, the basic person-to-person video call is not available for EDGE networks, which may limit the service's appeal to operators pursuing joint 3G/EDGE deployment strategies. Lack of EDGE support will mean a lower user base for the service, which limits the value of the service to users as they will be able to contact fewer people.

Some users are also reluctant to subscribe to the service because of privacy concerns—they often do not like the idea of their face being seen on somebody else's terminal first thing in the morning (Lemay, 2006).

Video calling is real time and therefore dependant on the quality of service with demands on the bandwidth and consistent packet interchange without interference, congestion, or static. In reality, the technology of today still does not seem to have the quality of service for video calling because it is directly dependent on real-time transfer of information (UMTS World, 2003).

The Future of 3G

The ways of doing business is changing the usage of IT and communication departments of every business amalgamating into one unit unlike never before, with the IT sector being the driving force for the benchmarks set by communication demand.

The mobile phone is ever changing, providing the user with one key characteristic in today's timeless world—convenience. With the mobile phone integrated into our day-to-day life in more ways than one, the introduction of IP into the communication world will transform its existence into a compact and user-friendly "computer". This will bring in the introduction of 4G.

Fourth Generation (4G) Cellular Mobile Networks

Third generation (3G) mobile networks have given emphasis on the development of a new network standard and hardware such as IMT2000, and the strategy to dynamically allocate frequency spectrum. The major hindrances for the introduction of 3G are the development of new technologies, and how to take away all frequencies from all countries and put them in a common pool. These problems are the main causes for the delay in the development of 3G networks. In spite of these, 3G can provide 2Mbps into a limited coverage area, that is, up to the macro (urban) cells. To solve these problems and to provide high-speed transmission and wider coverage, fourth generation mobile networks have been targeted to develop as IP-core heterogeneous networks. It is expected to be commercially launched by around 2010. There are a number of characteristics of 4G networks from users', operators', and service providers' point of views. From a user's point of view, 4G networks should have the following characteristics (Hui, 2003):

- The services of any networks ranging from body area to global communication networks including ad-hoc, sensors, and intelligent home appliances, should be available at anytime-anywhere irrespective of users' geographical locations and terminal devices. For example, users located anywhere in the world want an instantaneous access to their required service(s) provided by a number of

networks with their existing terminals. This is possible since 4G networks are IP-core.

- High transmission of data including 3D enhanced real-time audio and video information with low transmission cost. The demands of users for multimedia data are rapidly increasing every day. However, concomitantly, users want affordable and lower usage cost. The multimedia information, especially audio and video, requires a high bandwidth, which inherently incurs a high transmission cost since the Internet provides high transmission services with lower cost. To address this catch-22, 4G networks have taken IP as a core or backbone.
- Personalised and integrated services: Personalised services mean all types of users from different geographical locations and professions are able to receive the services they want in their customised ways using their available terminal devices. Integrated services mean users can simultaneously access multiple services from a number of service providers or networks using their same terminal devices. An example of integrated services is when a particular user with the same terminal device can concurrently browse the Internet using GPRS, make a phone call to their friend using GSM and see cricket live on television, or watch a program using W-CDMA networks (Hui & Yeung, 2003).

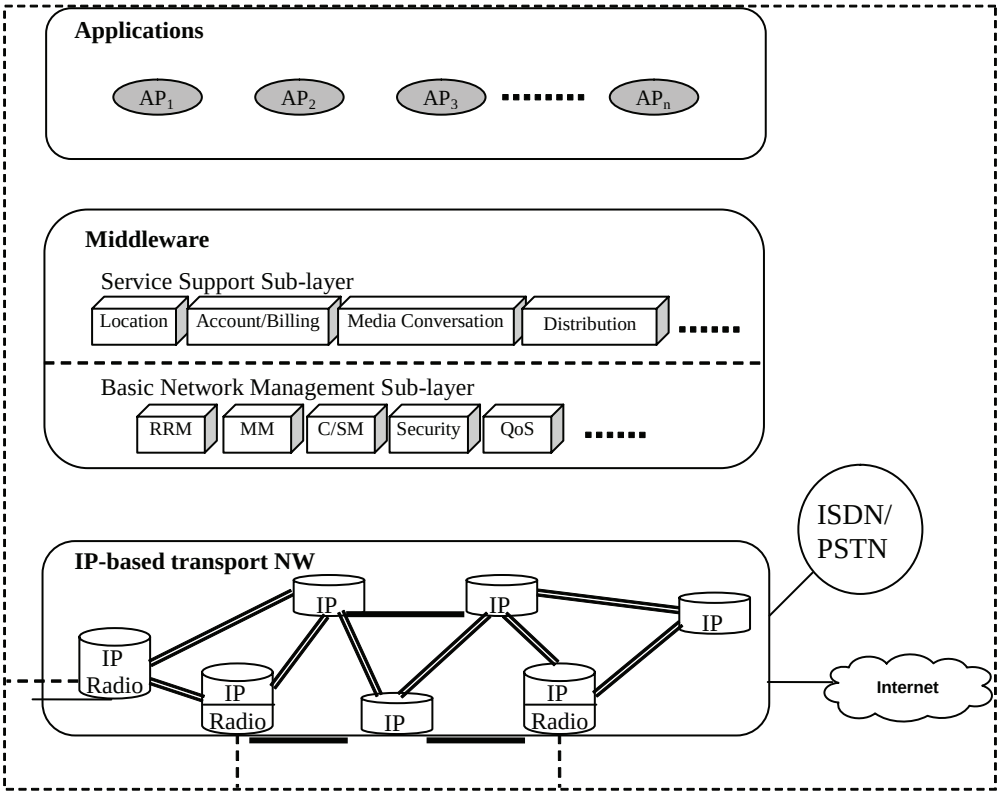
A wide variety of 4G network architectures has been developed, which can be classified into the following two approaches:

1. Interworking: For the interworking among different networks such as ad-hoc, wireless LAN, cellular, digital video broadcast-terrestrial (DVB-T), digital audio broadcast (DAB), satellite, and so on, new standard radio interfaces and technologies for 4G will be developed so they can provide seamless access and services to users. An example of 4G mobile networks architecture based on the interworking approach is presented in Muñoz and Rubio (2004), which comprises personal level (e.g., ad-hoc, wireless PAN), local level (e.g., Wireless LANs), national level (e.g., UMTS, GSM, and GPRS), regional level (e.g., DAB and DVB-T), and global coverage (e.g., satellite) networks.
2. Integration: During the development of 3G technologies, there are number of technologies which are being developed separately, such as Bluetooth, wireless local area networks (WLAN—IEEE 802.11), high performance radio LAN (HIPERLAN—IEEE 802.11a), and IMT2000. Each of these technologies has its own advantages and disadvantages. There is no such single technology which can provide the facilities of all these technologies. Therefore, researchers are giving emphasis on the development of 4G architecture by seamlessly integrating these existing technologies. A representative example of these architectures is presented in Figure 6 (Ruiz, 2002).

All existing wireless networks starting from wireless personal area networks (PANs) to global satellite networks are to be connected to an IP-based backbone in a hierarchical order. An example of a hierarchy of wireless networks connected to an IP-based backbone is shown in Figure 7.

This hierarchical structure confirms that when a user is connected to a personal network layer (e.g., in an office, airport, shopping mall, hotel, visitor centre, or train station), the user will receive a high bit rate transmission. In comparison, when a user moves in the upper layers (e.g., highway, beach, and remote area), the user will automatically be connected to the upper layer networks (e.g., WLAN, 2G and 3G cellular, DAB and DVB-T, and satellite). For mobility management, this structure supports both horizontal and vertical handoffs. When a mobile terminal moves through the same network layer, a horizontal handoff

Figure 6. General architecture of the IP-based 4G network platform based on integration approach



RRM= Radio resource management *MM: Mobility management*

C/SM: Connection/session management

takes place. If a mobile terminal moves across the network layers, a vertical handoff takes place. To meet the demands from users, operators, and service provides, and to seamlessly integrate the available network technologies, 4G networks possess a number of research challenges considering the aspects of mobile terminals, network systems, and mobile services that are described as in the following sections (Hui & Yeung, 2003; Ray, 2006).

Key Challenges of Mobile Terminals

Multimode Mobile Terminals

Single mobile terminals should be capable of accessing services from different mobile network technologies by automatically reconfiguring itself with them. The design of such a mobile terminal poses a number of problems related to high cost and power consumption, limitation in terminal size, and interpretational capability with downgraded technologies. An approach based on software radio has been proposed to adapt a mobile terminal with the interfaces of various network technologies (Buracchini, 2000).

Automatic Network System Discovery

A mobile terminal should be able to automatically detect its surrounding available wireless networks by receiving the signals broadcasted by them. For this, the multimodal terminal needs to scan the signal quickly as well as accurately from a number

of diverse networks comprising different access protocols. This makes the job difficult and challenging. To address this issue, a software radio is suggested to use in scanning the signals sent from heterogeneous networks.

Automatic Selection of a Network System

Multimode mobile terminal is required to automatically select a suitable wireless network for each service. This may need an optimisation of each user requirement for each session based on available network resources, network service capability, and both application and network quality of service (QoS) parameters. While this can be a time-varying multivariable optimisation technique, it may involve complex intelligent processing of different network characteristics as alluded to earlier, including user profile and preference. Despite a number of solutions which have been introduced considering some of the previously-mentioned criteria, a number of issues are yet to be solved to select a suitable network (Hui & Yeung, 2003).

Key Challenges of a Network System

Terminal Mobility

Terminal mobility means the ability of a communication network to locate a mobile device anytime anywhere in the world, and hence route a call to that particular device irrespective of its network point of attachment. This system facility allows a mobile user to have a global roaming, that is, to roam the different geographical locations all over the world. For this, network systems need to have an efficient mobility management system, which includes both location and handoff managements. Location management mainly locates and updates the location of a mobile terminal, while handoff allows the continuation of a call when a call is in progress, regardless of the movement of a user or terminal across wireless networks. IP-based mobility management has been incorporated in mobile IPv6 (MIPv6) (Hui & Yeung, 2003).

In spite of its enhancement to reduce high handoff latency and packet loss, handoff management remains a fairly complex process due to an unacceptable degradation of QoS for real-time multimedia transmission, and not being able to measure the accurate handoff time. This is because, as shown in Figure 7, 4G requires both horizontal and vertical handoffs across wireless networks. To address this issue, researchers are investigating ways to develop an efficient handoff management scheme.

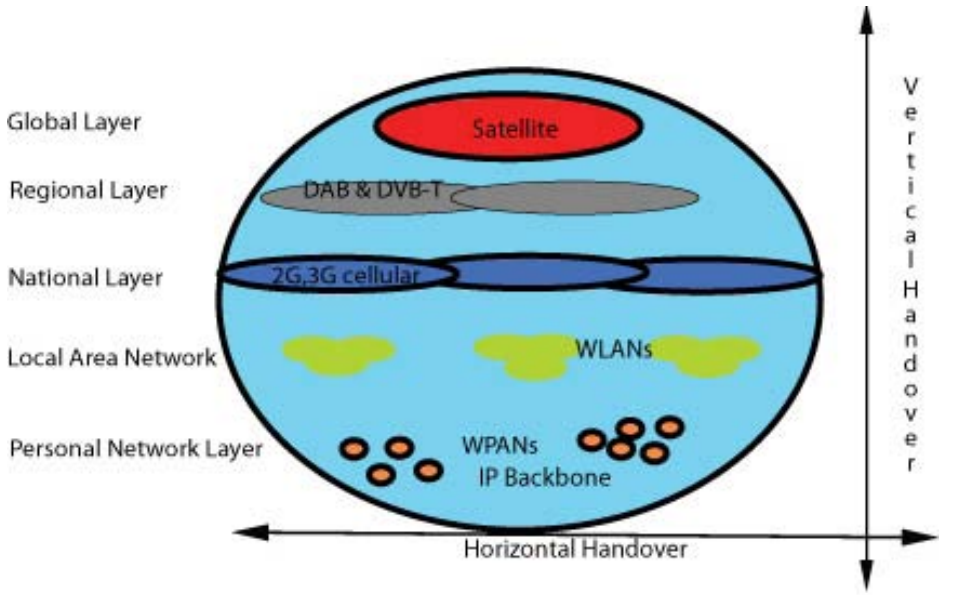
Adaptive and Light Weight Security System

Much of the existing mobile communication networks are mainly circuit switch based, that is, non-IP core. For this reason, the security standards which have been developed so far are mainly for non-IP based networks, and therefore are not directly applicable to IP-based mobile communication (4G) networks. In addition, to achieve the expected performance of QoS parameters, different standards have adopted different security schemes. For example, GSM has optimised its security for mainly voice delivery (Hui & Yeung, 2003). A particular security scheme is not suitable for 4G networks as it comprises a number of heterogeneous networks. Therefore, a light weight security system is needed to be developed, which can dynamically reconfigure itself considering the network and service types, operating environment, and user preference, and so forth.

Fault Tolerance and Network Survivability

Much effort has not been given to increase the fault tolerance and network survivability of wireless networks compared with wired networks. While the wireless networks are more vulnerable than wired networks due to its tree type hierarchical

Figure 7. Hierarchy of wireless networks connected to IP-based backbone



structure, if damage occurs in a particular level, it may affect the whole level including all of its underlying levels. This becomes more serious for 4G networks because it consists of heterogeneous multiple tree types networks (Hui & Yeung, 2003).

Key Challenges of Mobile Services

Integrated and Intelligent Billing System

As mentioned in the beginning of this section, users' demand for 4G mobile communication networks to access any service from any network using any mobile device anywhere in the world. This enforces the system to keep record of who uses what service from which service provider, and its related rate information. Each service provided by a particular provider has its own charging scheme with negotiated QoS parameters. For example, in Australia, charges for 3G mobile cricket are different from that of browsing the Internet through a mobile device. Even for the same service, the billing scheme varies from provider to provider. For example for a particular service, namely the Internet, some service providers offer a flat rate, while others charge as per usage, such as time and data. In addition to these, some providers have introduced their charging scheme based on usage time zone, such as peak, non-peak, and free time. These factors make the billing system challenging and complicated, which is difficult enough for users to record and pay all bills they received from each service provider separately. This enforces users to demand a unified single billing system. However, to produce a single bill using all usage history requires an intelligent broker service (Hui & Yeung, 2003). In order to produce such a unified billing system, researchers are now delving into all possibilities.

Personal Mobility

Personal mobility is one aspect of mobility management and global roaming. Personal mobility refers to the capability of a user to access any service from any network using any mobile terminal regardless of his network attachment point. Thus, this important characteristic of mobility management enables a user to receive a

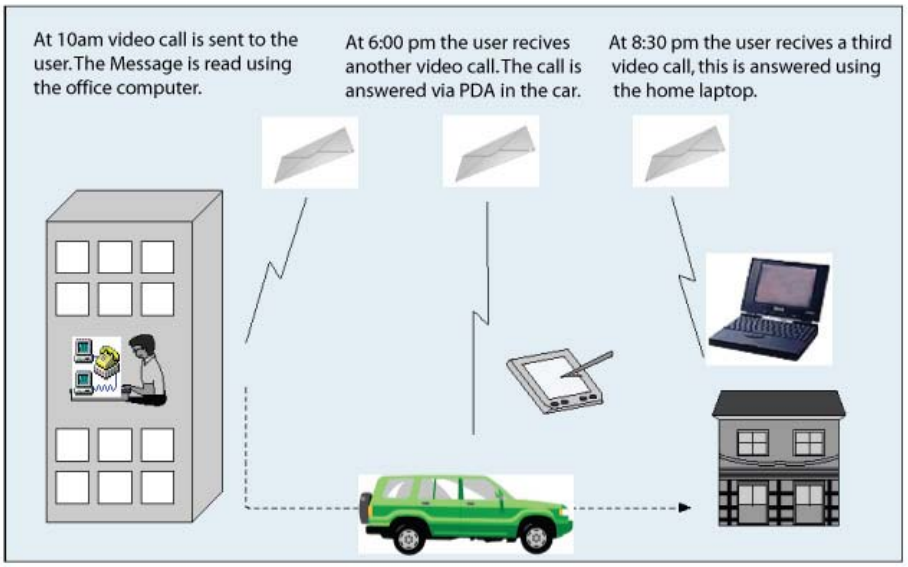
personalised service through his profile, that is, his preferred location, choice, and mobile devices. An example of a personalised service for a particular user, namely Mary for her video messages, is shown in Figure 8 (Hui & Yeung, 2003).

For this, a user is accessing her personalised video call at different times, and in different operating environments. Researchers have introduced a number of schemes for personal mobility management. Most of them are mainly based on mobile agent.

SUMMARY

The birth of mobile communication is mainly to fulfil people's expectation to communicate with anyone, anytime, anywhere. The transmission of multimedia through mobile communication networks and the competition to present equivalent services of wired communication networks is widely in demand. The elements of multimedia are text, image, audio, and video. The audio includes voice and music. The first generation mobile communication networks were analog, and mainly for voice services. There was no security scheme. The communication speed was also very low and covered a limited distance, and a limited number of users. To address these issues, there was a birth of 2G mobile communication standards which are in digital transmission. In addition, to afford higher voice transmission rates, 2G standards also provide SMS and data services. SMS services have been very popular with young people all over the world. Since 2G standards are unable to support e-mail and Internet services, some 2G standards have been enhanced to 2.5G to increase the transmission speed, as well as to afford those services. The maximum transmission rate supported by 2.5G is 384 kbps, which is not sufficient for application services that involve audio and video transmission such as mobile TV and video conferencing. For this, 3G generation mobile communication has recently appeared in the market to provide broadband communication, that is, up to 2Mbps transmission. This transmission rate is not sufficient for real-time multimedia information communication.

Figure 8. Personal communication using a personalised video message



Furthermore, the future of 3G is not clear enough because of the need to develop new standards and hardware, as well as the difficulties to develop a unified frequency pool. However, there are other technologies that are available such as Bluetooth, Wireless LAN, HyperLAN, GPRS, and UMT2000. Each system has their own merits that are worthy to consider for 4G. These factors, including the demand for personalised communication services from any service provider all over the world, 4G mobile communications standards are projected to develop and include existing cutting-edge technology, and future technologies. Since the different standards have different types of services, security and billing systems, it poses a huge key research challenge for mobile terminals, network systems, and services that are also briefly described in this chapter.

REFERENCES

Akyildiz, I. F., McNair, J., Ho, J., Uzunalioglu, H., & Wang, W. (1999). Mobility management in next-generation wireless systems. *Proceedings of the IEEE, 87*(8), 1349-1351.

Andersson, C. (2001). *GPRS and 3G wireless applications.* NY: John Wiley & Sons, Inc.

Ariyoshi, M., Shima, T., Han, J., Karlsson, J., & Urabe, K. (2002). Effect of forward-backward filtering channel estimation in W-CDMA multi-stage parallel interference cancellation receiver. *IEICE Trans. Commun., E85-B*(10), 1898-1905.

Bates, R. J. (1995). *Wireless networked communications: Concepts, technology, and implementations.* NY; Singapore: McGraw-Hill, Inc.

Buracchini, E. (2000). The software radio concept. *IEEE Communication Magazine, 38,* 138-143.

Casal, C. R., Schoute, F., & Prasad, R. (1999). A novel concept for fourth generation mobile multimedia communication. *IEEE 50th Vehicular Technology Conference (VTC 1999-Fall), 1,* Amsterdam, The Netherlands, September 19-22 (pp, 381-385). Piscataway, NJ: IEEE Service Center.

Chapman, N., & Chapman, J. (2000). *Digital multimedia*. London: John Wiley & Sons.

ePanorama. (2006). Retrieved February 2007, from http://www.epanorma.net/links/tele_mobile.html

Garfield, L. (2004). *Infosync: Reporting from the digital*. Retrieved May 2007, from http://www.infosyncworld.com/news/n/5048.html

Gartner. (2005). *Smart phones are favoured as thin clients for mobile workers.* Retrieved May 2007, from http://www.nokia.com/NOKIA_COM_1/About_Nokia/Press/White_Papers/pdf_files/Whitepaper_TheMythsofMobility.pdf

Hackett, S. (2001). *Aglie communication: m.Net Australia consortium wins federal funding for 3G/WLL test bed and applications development environment.* Retrieved May 2007, from http://www.agile.com.au/press/press-29-05-2001.htm

Halsall, F. (2001). *Multimedia communications: Applications, networks, protocols and standards.* Harlow, England; NY: Addision-Wesley.

Hui, S. Y., & Yeung, K. H. (2003). Challenges in the migration to 4G mobile systems. *IEEE Communication Magazine, 41*(12), 54-59.

IFC. (2005). *IFC: Universal mobile telecommunications system (UMTS) proto.* Retrieved May 2007, from http://www.iec.org/online/tutorials/umts/topic01.html

Le Maistre, R. (2002). *US to top 3G chart in 2010 European editor, Unstrung.* Retrieved May 2007, from http://www.unstrung.com/document.asp?doc_id=24887

Lemay, R. (2006). *Perth, Adelaide get optus, Vodafone 3g.* Retrieved May 2007, from http://www.zdnet.com.au/news/communications/soa/

perth-adelaide-get-optus-vodafone-3g/0,130061791,139261668,00.htm

McCullough, J. (2004). *185 wireless secrets: Unleash the power of PDAs cell phones, and wireless networks.* Indianapolis, IN: Wiley Publishing, Inc.

Mobile. (2007). *Mobile computing.* Retrieved May 2007, from http://searchmobilecomputing.techtarget.com/sDefinition/0,,sid40_gci506042,00.html

Muñoz, M., & Rubio, C. G. (2004). A new model for service and application convergence in B3G/4G networks. *IEEE Wireless Communications, 35*(5), 539-549.

Nokia. (2006). *State of workforce mobility.* Retrieved May 2007, from http://www.nokia.com/NOKIA_COM_1/About_Nokia/Press/White_Papers/pdf_files/Whitepaper_TheMythsofMobility.pdf

Ohya, T., & Miki, T. (2005). Enhanced-reality multimedia communications for 4G mobile networks. *1st International Conference on Multimedia Services Access Networks (MSAN '05),* Orlando, FL, June 13-15 (pp. 69-72). Piscataway, NJ: IEEE Service Center.

Orr, E. (2004). *3G-324M helps 3G live up to its potential.* Retrieved May 2007, from http://www.wsdmag.com/Articles/Print.cfm?ArticleID=7742

Rao, M., & Mendoza, L. (Eds.) (2005). *Asia unplugged: The wireless and mobile media boom in the Asia-Pacific.* New Delhi: Response Books (A Division of Sage Publications).

Ray, S. K. (2006). Fourth generation (4G) networks: Roadmap-migration to the future. *IETE Technical Review, 23,* 253-265.

Ruiz, P. M. (2002). *Beyond 3G: Fourth generation wireless networks.* Retrieved May 2007, from http://internetng.dit.upm.es/ponencias-jing/2002/ruiz/ruiz.PDF

Salzman, M., Palen, L., & Harper, R. (2001). *Mobile communications: Understanding users, adoption and design.* CHI 2001, Seattle, WA, March 31-April 5.

Sawada, M., Tani, N., Miki, M., & Maruyama, Y. (1998). Advanced mobile multimedia services and applied network techniques. *IEEE International Conference on Universal Personal Communications, 1,* 79-85.

UMTS World. (2003). *3G applications.* Retrieved May 2007, from http://www.umtsworld.com/applications/applications.htm

Wesel, E. K. (1998). *Wireless multimedia communications: Networking video, voice, and data.* Addison Wesley.

Wikipedia. (2006). *Wikipedia, the free encyclopedia.* Retrieved October 2006, from http://en.wikipedia.org/wiki/

This work was previously published in Mobile Multimedia Communications: Concepts, Applications, and Challenges, edited by G. Karmakar and L. Dooley, pp. 1-23, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.11
Discovering Multimedia Services and Contents in Mobile Environments

Zhou Wang
Fraunhofer Integrated Publication and Information Systems Institute (IPSI), Germany

Hend Koubaa
Norwegian University of Science and Technology (NTNU), Norway

ABSTRACT

Accessing multimedia services from portable devices in nomadic environments is of increasing interest for mobile users. Service discovery mechanisms help mobile users freely and efficiently locating multimedia services they want. The chapter first provides an introduction to the topic service discovery and content location in mobile environments, including background and problems to be solved. Then, the chapter presents typical architectures and technologies of service discovery in infrastructure-based mobile environments, covering both emerging industry standards and advances in the research world. Their advantages and limitations, as well as open issues are discussed, too. Finally, the approaches for content location in mobile ad hoc networks are described in detail. The strengths and limitations of these approaches with regard to mobile multimedia services are analyzed.

INTRODUCTION

Recently, the advances in mobile networks and increased use of portable devices deeply influenced the development of multimedia services. Mobile multimedia services enable users to access multimedia services and contents from portable devices, such as laptops, PDAs, and even mobile phones, at anytime from anywhere. Various new applications, that would use multimedia services on portable devices from both the fixed network backbone and peer mobile devices in its proximity, are being developed, ranging from entertainment and information services to business applications for M-Commerce, fleet management, and disaster management.

However, to make mobile multimedia services become an everyday reality, some kinds of service infrastructures have to be provided or enhanced, in order to let multimedia services and contents on the network be discovered and utilized, and simultaneously allow mobile users to search and request services according to their own needs, independently of the physical places they are visiting and the underlying host platforms they are using. Particularly, with the explosive growth of multimedia services available in the Internet, automatic service discovery is gaining more and more significance for mobile users. In this chapter we focus on the issue of discovering and locating multimedia services and contents in mobile environments. After outlining necessary background knowledge, we will take an insight into mobile multimedia service discovery. Major service discovery architectures and approaches in infrastructure-based networks and in mobile ad hoc networks will be investigated. We present also a detailed analysis of their strengths and limitations with regard to mobile multimedia services.

DISCOVERING MOBILE MULTIMEDIA SERVICES AND CONTENTS IN INFRASTRUCTURE-BASED ENVIRONMENTS

Overview

In order to use various multimedia services on the network, the first necessary step is to find the exact address of service providers that implement the service. In most cases, end users might only know what kind of service (service type) and some service characteristics (e.g., data format, cost) they want, but without having the server address. Currently, browsing is one often-used method to locate relevant information. As the number and diversities of services on the network grow, mobile users may be overwhelmed by the sheer volume of available information, particularly in an unacquainted environment. On the other side, user mobility presents new challenges for service access. Mobility means that users probably change their geographic locations frequently. Consequently, services available to users will appear or disappear dynamically while users move here and there. Moreover, mobile users are often interested in the services, (e.g., malls, restaurants) in the close proximity of his or her current place. Therefore, unlike classical distributed environments where location is often kept transparent, applications often need to dynamically obtain information that is relevant to their current location. The service search procedure should be customized according to user's context, (e.g., in terms of when (i.e., time) and where (i.e., location) a user is visiting).

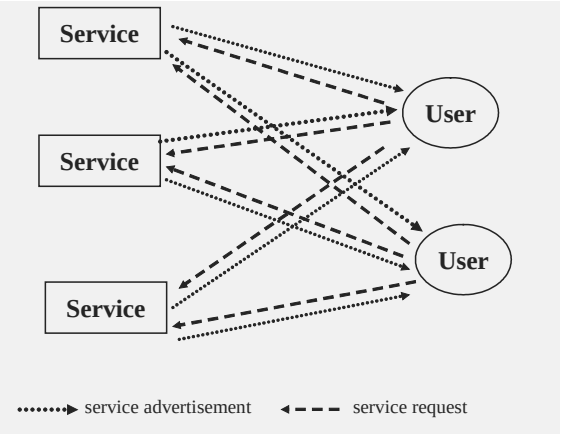
Since most current multimedia services are designed for stationary environments, they do not address these issues. Recently, a number of service discovery solutions are developed. These solutions range from hardware-based technologies such as Bluetooth SDP, to single protocols, (e.g., SLP and SDS) to frameworks such as UPnP and Jini. From architectural point of view, we observed three models are used to discover services in different network environments (Wang, 2003): the broadcast model, the centralized service directory model, and the distributed service directories model. Next, we will investigate these paradigms in detail.

Broadcast Model

The simplest architecture for service discovery is using broadcast to locate services and contents. The conceptual scheme of the broadcast model is depicted in Figure 1. In this model, clients and servers talk directly with each other through broadcast or multicast.

According to who initiates the announcement and who listens, two strategies are differentiated. The first strategy is the *pull strategy* where a

Figure 1. Broadcast model



client announces his requests, while all servers keep listening to requests. The servers that match the search criteria will send responses (using either unicast or multicast) to the client. The other strategy is the *push strategy*. The servers advertise themselves periodically. Clients who are interested in certain types of services listen to the service advertisements, and extract the appropriate information from service advertisements. Of course, hybrid strategies are applied by some approaches.

The **simple service discovery protocol** (**SSDP**) is one typical approach based on the broadcast model (Goland, Cai, Leach, Gu, & Albright, 1999). The SSDP builds upon HTTP and UDP-multicasting protocols, and employs a hybrid structure combining client announcement and service announcement. When a device is newly added to the network, it multicasts an "ssdp:alive" message to advertise its presence. Similarly, when a client wants to discover services, it multicasts a discovery message and awaits responses.

The broadcast model works well in small simple networks, such as home and small office. The primary advantage of such systems is that they need "zero" or little configuration and administration. Besides, they accommodate well to frequent service join/leave actions in a dynamic environment. However, they usually generate heavy network traffic due to broadcast, and thus have only minimal scalability.

In order to improve scalability and performance, an additional entity, service directory, is introduced. Two different models use the service directory: the centralized service discovery model and the distributed service directories model. Both models will be presented in the following sections.

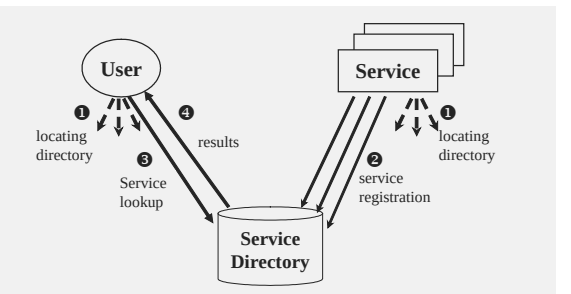
Centralized Service Directory Model

The conceptual scheme of the centralized directory model is shown in Figure 2. The service directory becomes the key component in the search discovery architecture, because it stores information about all available services.

The service discovery procedure consists usually of the following steps:

1. **Locating directory:** Either clients or servers should determine the address of the service directory before they utilize or advertise services. The directory could be located by manual configuration, by querying a well-known server, or through broadcast/multicast requests/replies.
2. **Service registration:** Before a service can be found by clients, it must be registered in the appropriate directory. A service provider

Figure 2. Centralized directory model



explicitly initiates a registration request to the directory, and the directory stores the service data in its database. The service description data include service type, service attributes, server address, etc.

3. **Service lookup:** As a client searches for a particular service, he describes his requirements, e.g. service type and desired characteristics, in a query request, and sends it to the directory.
4. **Searching:** The directory searches services in its database according to the criteria provided by the client. When services are found, the server addresses and other information of qualified services are sent back to the client.

The centralized directory model has been used by several service discovery approaches. In this section we will examine some of them.

Service Location Protocol (SLP)

The service location protocol (SLP) is an example of centralized directory-based solution, and is now an IETF standard (Guttman, Perkins, Veizades, & Day, 1999). The current version is SLP Version 2 (SLPv2). The SLP uses DHCP options, or UDP-based multicasting to locate the service directory (known as directory agent (DA)), without manual configuration on individual clients and services (known as user agents (UAs), service agents (SAs) respectively). A multicast convergence algorithm is adopted in SLP to enhance multicast reliability.

Service registration and lookup are performed through UDP-based unicast communication between UAs/SAs and DAs. In addition, SLP can operate without DAs. In this mode, SLP works in the same way as the broadcast model. A service in SLP is described with service type in the form of a character string, the version, the URL as server address, and a set of attribute definitions in the form of key-value pairs.

To improve performance and scalability, more DAs can be deployed in network. However, SLPv2 does not provide any synchronization mechanisms to keep DAs consistent, but leaves this responsibility to SAs which should register with each DA they detect. Recently, (Zhao & Guttman, 2000) proposed a mesh enhancement for DAs to share known services between one another. Each SA needs to register only with a single DA, and its registration is automatically propagated among DAs.

Generally, SLP is a flexible IP-based service discovery protocol which can operate in networks ranging from a single LAN to an enterprise network. However, it is intended to function within networks under cooperative administrative control, and thus does not scale for the Internet.

JINI

Sun's JINI provides a similar architecture as SLP for delivering services in a network (Sun Microsystems Inc., 2003), but it is tightly bound to the Java environment and needs Java Virtual Machine (JVM) support. The protocols in JINI are implemented as Java APIs. For this reason, the JINI client is not as lightweight as the SLP client. However, JINI is more than a discovery protocol. It provides further facilities for service invocation, for transaction, and for distributed events.

INS

Adjie-Winoto, Schwartz, Balakrishnan, and Lilley, (1999) proposes a resource discovery system named *intentional naming system* (INS). The main idea is that resources or services are named using an ordered list of attribute-value pairs. Since service characteristics can be described by the service name itself, the service discovery procedure is equal to name resolving which is accompanied by the *intentional name resolver* (INR). The INR is actually a service directory that holds the global knowledge about names in

the whole network. INS is different from other naming services (e.g., DNS), in that the name describes service attributes and values, rather than simple network locations of objects.

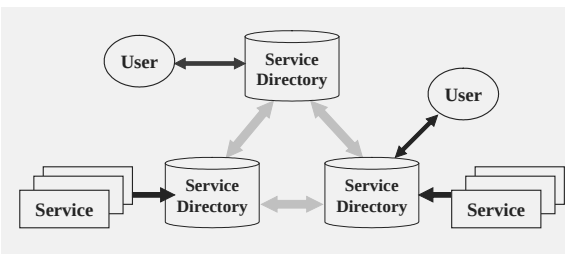
In conclusion, most centralized directory-based architectures have been designed for local networks or enterprise-wide networks which are under a common administration. The primary issue for these systems is scalability. As the number of services and clients increases, a centralized directory, even replicated, will not be feasible to accommodate a large number of registrations and lookups. In this context, the distributed repositories model has been suggested.

Distributed Service Directories Model

In the distributed directories model, the whole service domain is split into partitions, possibly according to organizational boundary, network topology, geographic locations etc. In each partition, there are one or more directories. The conceptual scheme of the distributed directories model is shown in Figure 3. The distributed directories model is different from the centralized directory model in that no directory has a complete global view of services available in the entire domain. Each directory holds only a collection of services in its partition, and is responsible for interaction with clients and services in the partition.

The service registration and query submission in the distributed model remain similar to that in the centralized directory model. But the service search operation becomes more complicated. If required services can be found by local directories, the discovery procedure is akin to that in the centralized directory model. But if not, the directories in other partitions should be asked, to ensure that a client can discover any service offers in the entire domain.

Figure 3. Distributed directories model



The directories in this model are organized in some way to achieve cooperation. As stated in (Wang, 2003), the directories can be organized in a hierarchy structure or in a mesh structure. While in the hierarchy structure there is a "belong to" relationship between directly connected directories, directories in the mesh architecture are organized in a flat interconnected form without hierarchy. The interconnection structure might have strong implications on query routing. In the hierarchy structure queries are passed along the hierarchy, either upward or downward, thus the routing path is inherently loop free. But the rigid hierarchy obstructs to shortcut the routing path in some cases. On the other hand, the mesh structure is advantageous for optimizing the routing path, but might rely on some mechanisms to avoid loop circles or repeated queries.

A typical example of distributed directories-based architecture is **service discovery service (SDS)**, developed in Berkeley (Hodes, Czerwinski, Zhao, Joseph, & Katz, 2002). The SDS is based on the hierarchy model which is maintained by periodic "heartbeat" messages between parent and child nodes. Each SDS server pushes service announcements to its parent. By this means, each SDS server gathers a complete view of all services present in its underlying tree. The significant feature of SDS is the hierarchical structure with lossy aggregation to achieve better scalability and reachability. The SDS server applies multiple hash functions (e.g., MD5), to various subsets of tags in the service description and uses the results to set bits in a fixed-size bit vector. The parent node ORs all bit vectors from its children to summarize available services in the underlying tree.

The hierarchical structure with lossy aggregation helps SDS to reach better scalability, while ensuring users to be able to discover all services on all servers. However, the SDS is more favorable for applying in stationary network environments since it requires additional overheads to maintain the hierarchical structure and to propagate index updates. If services change attributes rapidly or join/leave frequently, it will generate too much communication burden. Moreover, the OR-operation during aggregation may cause "false positive" answers in query routing. Although it does not sacrifice correctness, it will lead to unneeded additional query forwarding.

The media gateway discovery protocol (MEGADIP) is developed especially for discovering media gateways that act as proxy for transforming or caching data between media source and end users (Xu, Nahrstedt, & Wichadakul, 2000). In MEGADIP the discovery procedure starts from the local directory, and forwards the query to directories along the routing path of the network layer between media source and destination. This idea is driven by the heuristics that a media gateway on or close to the end-to-end path is likely to find more bandwidth and/or to incur smaller end-to-end delay.

Other Issues in Service Discovery

The architectural models and various approaches presented above solved the service discovery problem to some extent. However, in order to let users comfortably and effectively locate mobile multimedia services and contents, there are still some issues to be addressed. From our point of view, interoperability, asynchronous service discovery, and semantic service discovery are the most important.

Interoperability

As previously stated, a number of service discovery approaches have been proposed. Despite that most of them provide similar functionality, namely automatically discovering services based on service characteristics, they have different features and are not compatible with each other. This incompatibility is one of the biggest obstacles for mobile users to really benefit from service discovery. From our point of view, it is more useful to make different approaches interoperable, than to design a new protocol to cover functionalities of existing protocols. So far, some solutions have been proposed to bridge service discovery mechanisms, but they are limited to pair-wise bridges, such as JINI to SLP (Guttman & Kempf, 1999). Authors in Friday, Davies, and Catterall (2001) proposed a general solution on a modified form of the Structured Query Language (SQL). However, no implementation details are presented in the paper. More generally, Wang and Seitz (2002) addressed this issue by providing an intermediary layer between mobile users and underlying service discovery protocols. The intermediary layer on the one hand provides clients with a general consistent view of service configuration and a universal means to formulate search requests, on the other hand is capable of talking with various types of service discovery protocols and handling service requests from users.

Asynchronous Service Discovery

Apart from the heterogeneous environments, most of the existing approaches rarely take the issues of thin client and poor wireless link into consideration. For example, synchronous operation is one of the intrinsic natures of most existing service discovery approaches, such as SLP, Jini, and SDS. Although synchronous operation simplifies protocol and application design, it is fastidious for mobile environments. The unexpected but frequent disconnections and possible long delay of wireless link greatly influence the usefulness and efficiency of synchronous calls. To relax the communication restrains in wireless environments, (Wang & Seitz, 2002) proposed

in their CHAPLET system an approach to achieve asynchronous service discovery by adopting mobile agents. The asynchronous service discovery allows mobile users to submit a service request, without having to wait for results, nor continuously keeping the permanently active connection in the process of service discovery.

Semantic Service Discovery

Most existing service discovery approaches support only syntactic-level searching (i.e., based on attribute comparison and exact value matching). However, it is often insufficient to represent a broad range of multimedia services in real world, and lacks of capability to apply inexact matching rules. Therefore, there is need to discover services in a semantic manner. Chakraborty, Perich, Avancha, and Joshi (2001) proposes in the DReggie project to use the features of DAML to reason about the capabilities and functionality of different services. They designed a DAML-based language to describe service functionality and capability, enhanced the Jini Lookup Service to enable semantic matching process, and provided a reasoning engine based on Prolog. Yang (2001) presents a centralized directory-based framework for semantic service discovery. However, the semantic-based service discovery is still in its infancy. To promote wide development of semantic service discovery, more research efforts should be devoted.

DISCOVERING MULTIMEDIA SERVICES AND CONTENTS IN AD HOC ENVIRONMENTS

Overview

There are two well-known basic variants of mobile communication networks: infrastructure-based networks and ad hoc networks. Mobility support described in the previous sections relies on the existence of some infrastructure. A mobile node in the infrastructure-based networks communicates with other nodes through the access points which act as bridge to other mobile nodes or wired networks. Normally, there is no direct communication between mobile nodes. Compared to infrastructure-based networks, ad hoc networks do not need any infrastructure to work. Nodes in ad hoc networks can communicate if they can reach each other directly or if intermediate nodes can forward the message. In recent years, mobile ad hoc networks are gaining more and more interest both in research and industry. In this section we will present some typical approaches that enable discover and locate mobile multimedia services and contents in ad hoc environments. First we present broadcast-based approaches, and then the geographic service location approach is discussed. Next, a cluster-based approach is introduced. Finally, we present a new service or content location solution that addresses the scalability problem in multi-hop ad hoc networks.

Broadcast-Based Approaches

Considering the fact that no infrastructure is available in ad hoc environments, service directory-based solutions are unusable for service discovery in ad hoc networks. Instead, assuming that network supports broadcasting, service discovery through broadcast is one of most widely adopted solutions. Two broadcast-based approaches are possible: (1) broadcasting client requests and (2) broadcasting service announcements. In the first approach, clients broadcast their requests to all the nodes in the ad hoc network. Servers hosting requested services reply back to the clients. In the second approach, servers broadcast their services to all the nodes in the network. Each client is thus informed about the location of every service in the ad hoc network. Since these both approaches are mainly based on broadcasting, their efficiency strongly depends on the broadcast efficiency. The service location

problem in that context can be reduced to the broadcast problem in ad hoc networks. For this reason, in the following, we present a summary of proposed approaches for broadcasting in ad hoc networks. These broadcast approaches are not designed specifically for service location but we believe that a broadcast-based service location protocol has to be informed about how broadcast is carried out. This will help in deploying a cross layer-based service location protocol.

The broadcast techniques can be categorized into four families: Williams and Camp (2002), simple flooding, Jetcheva, Hu, Maltz, and Johnson (2001), probabilistic broadcast, Tseng, Ni, Chen, and Sheu (1999), location-based broadcast, and neighbor information broadcast, Lim and Kim (2000) and Peng and Lu (2000). Flooding represents a simple mechanism that can be deployed in mobile ad hoc networks. Using flooding, a node having a packet to be broadcasted sends this packet to his neighbors who have to retransmit it to their own neighbors. Every node receiving the packet for the first time has to retransmit it. To reduce the number of transmissions used in broadcasting, other broadcast approaches are proposed. The probabilistic broadcast is similar to flooding except that nodes have to retransmit the broadcast packet with a predetermined probability. Randomly choosing the nodes that have to retransmit can improve the bandwidth use without influencing the reachability. In the case of location-based broadcast techniques, a node x retransmits the broadcast packet received from a node y only if the distance between x and y exceeds a specific threshold.

The information on the neighborhood can also be used to minimize the number of nodes participating in the broadcast packet retransmission. Lim and Kim (2000) uses the information about the one hop neighborhoods. Node A, receiving a broadcast packet from node B, compares its neighbors to those of B. It retransmits the broadcast packet only if there are new neighbors that will be covered and that will receive the broadcast packet. Other broadcast protocols are based on the 2 hop neighborhood information. The protocol used in Peng and Lu (2000) is similar to the one proposed in Lim and Kim (2000). The difference is that in Lim and Kim (2000) the neighborhood information is sent within HELLO packets, whereas in Peng and Lu (2000), the neighborhood information is enclosed within the broadcast packet.

The study carried out in Williams and Camp (2002) showed that the probabilistic and location broadcast protocols are not scalable in terms of the number of broadcast packet retransmissions. The neighborhood-based broadcast techniques perform better by minimizing the number of nodes participating to the broadcast packet retransmission. The most significant disadvantage of these protocols is that they are sensitive to mobility.

Geographic Service Location Approaches

A more interesting service location approach than broadcasting the whole network is to restrict broadcasting to certain regions. These regions can be delimited on the basis of predefined trajectories. In fact, recently, geometric trajectories are proposed to be used for routing (Nath & Niculeson, 2003) and content location in location-aware ad hoc networks (Aydin & Shen, 2002; Tchakarov & Vaidya, 2004). Aydin and Shen (2002) and Tchakarov and Vaidya (2004) are closely related where content advertisements and queries are propagated along four geographical directions based on the physical location information of the nodes. At the intersection point of the advertising and query trajectories the queries will be resolved. Moreover, Tchakarov and Vaidya (2004) improves the performance by suppressing update messages from duplicate resources. However, basically they still rely on propagating advertisements and queries through the network.

Cluster-Based Solutions

Besides enhancements in broadcast, clustering can also be used to improve the performance of service discovery in mobile ad hoc networks. An interesting cluster-based service location approach designed for ad hoc networks is proposed in Koubaa and Fleury (2001) and Koubaa (2003). The proposed approach involves four phases: (1) the servers providing services are organized within clusters by using a clustering protocol. The cluster-heads, elected on the basis of an election protocol, have the role of registering the addresses of the servers in their neighborhoods (clusters). (2) A reactive multicast structure gathering the cluster-heads to which participate the cluster-heads of the created clusters is formed at the application layer. Each client or a server in the network is either a part of this structure or one hop away from at least one of the multicast structure members. (3) Clients send their request inside this multicast structure. (4) An aggregation protocol is used to send the replies of the cluster-heads within the multicast structure. The aim of the aggregation protocol is to avoid using different unicast paths for reply transmission by using the shared paths of the multicast structure.

A study comparing broadcast approaches to the cluster-based approach is carried out in Koubaa and Fleury (2002). This comparison study showed that clustering reduces the overhead needed for clients to send their requests and for servers to send back their replies. This reduction is noticeable when we increase the number of clients, the number of servers, and the number of nodes in the ad hoc network. The multicast structure used in Koubaa (2003) consists of a mesh structure which is more robust than a tree structure. The density of the mesh structure is dynamically adapted to the number of clients using it. The key idea of this dynamic density mesh structure is that the maintaining of the mesh is restricted to some clients called effective clients. Indeed, when the network is dense or the number of clients is high there is no need that all clients participate the multicast structure maintaining. This new mesh structuring approach is compared to ODMRP (Koubaa, 2003) where all the multicast users participate in the mesh maintaining. The comparison study showed that the proposed dynamic density mesh is more efficient than ODMRP. Compared to the tree-based multicast structure, the mesh-based multicast structure shows better server reply reachability performance but using more bandwidth.

Scalability Issue in Service Location

Currently it is well known that ad hoc networks are not scalable due to their limited capacity. The scalability problem is mainly related to the specific characteristics of the radio medium limiting the effective ad hoc network capacity. Even though, we think that designing specific solutions for scalable networks can help us at defining how much scalable is an ad hoc network. In the context of service location, authors in Koubaa and Wang (2004) state the problem of scalable service location in ad hoc networks and propose a new solution inspired by peer-to-peer networks called HCLP (hybrid content location protocol). The main technical highlights in approaching this goal include: (1) the hash function for relating content to zone, (2) recursive network decomposition and recomposition, and (3) content dissemination and location-based on geographical properties.

The hashing technique is used in HCLP both for disseminating and locating contents. But unlike the approaches in peer-to-peer systems where the content is mapped to a unique node, the hash function in HCLP maps the content to a certain zone of the network. A zone means in HCLP a certain geographical area in the network. The first reason for mapping content into zone, i.e. a subset of nodes, instead of an individual node, is mainly due to the fact that it could be expensive in radio mobile environments to maintain a predefined rigid structure between nodes for routing adver-

tisements and queries. For example, in Stoica, Morris, Karger, Kaashoek, and Balakrishnan (2001), each joining and leaving of nodes has to lead to an adjustment of the Chord ring. Moreover, the fact that the routing in ad hoc networks is far less efficient and less robust than in fixed networks makes the adjustments more costly if there is node movement. The second reason for relating content to zone is that it is more robust to host a content within many nodes inside a zone than to host it within an individual node.

The underlying idea of network decomposition in HCLP is to achieve load distribution by maintaining the zone structure. It is well known that if the number of the nodes and contents in an unstructured and decentralized zone is beyond a certain limit, the network overhead related to content advertisement/location would become unsatisfactory. Therefore, to ensure a favorable performance and to achieve a better load distribution in HCLP, a zone could be divided into sub-zones recursively if the cost related to content advertisement/location using unstructured approaches in the zone exceeds a certain threshold.

To enable network decomposition in different zones a protocol is deployed to make it possible to nodes on the perimeter of the network exchanging their geographical locations. This will help estimating the position of the centre of the network. Knowing the locations of the nodes on the perimeter and the location of the network centre, a simple decomposition of the network into four zones is used. Each of these zones can also be decomposed again into four zones, etc.

In HCLP, for disseminating or locating a content in the network, a user first sends out its announcement or query request along one of four geographical directions (north, south, east, and west) based on geographic routing. In a dense network, the announcement or the request will then be caught on the routing path by a node that knows the central region of the network, in the worst case by a perimeter node on the network boundary. This node will then redirect the request into the direction of the central region, again by geographic routing. The node that belongs to the central region and receives this query message has the responsibility to decide whether to resolve the request directly within the zone or whether to redirect the request to the next level of the zone hierarchy, until the content is discovered.

Such a content dissemination and location scheme works completely decentralized. Moreover, only a small portion of nodes is involved in routing and resolving advertisement or query messages. Because not all nodes are necessary for maintaining routing information nor a global knowledge of the whole network is required, HCLP can be expected to be well scalable to large ad hoc networks.

CONCLUSION

The prevalence of portable devices and wide deployment of easily accessible mobile networks promote the usage of mobile multimedia services. In order to facilitate effectively and efficiently discovering desirable mobile multimedia services and contents, many research efforts have been done. In this chapter, we discussed existing and ongoing research work in the service discovery field both for infrastructure-based mobile networks and mobile ad hoc networks. We introduced three main architectural models and related approaches for service discovery in infrastructure networks, and pointed out some emerging trends. For discovering services and contents in ad hoc networks, we presented and compared proposed approaches based on either broadcast or cluster, and discussed the scalability issue in detail. We believe that service discovery will play an important role for successful development and deployment of mobile multimedia services.

REFERENCES

Adjie-Winoto, W., Schwartz, E., Balakrishnan, H., & Lilley, J. (1999). The design and implementation of an intentional naming system. In *Proceedings of the 17th ACM Symposium on Operating Systems Principles (SOSP ´99).*

Aydin, I., & Shen, C. (2002, October). *Facilitating match-making service in ad hoc and sensor networks using pseudo quorum*. In the 11th IEEE International Conference on Computer Communications and Networks (ICCCN).

Chakraborty, D., Perich, F., Avancha, S., & Joshi, A. (2001, October). *DReggie: Semantic service discovery for m-commerce applications.* In the Workshop on Reliable and Secure Applications in Mobile Environment, in Conjunction with 20th Symposium on Reliable Distributed Systems (SRDS).

Friday, A., Davies, N., & Catterall, E. (2001, May). Supporting service discovery, querying, and interaction in ubiquitous computing environments. In *Proceedings of the 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access,* Santa Barbara, CA (pp. 7-13).

Goland, Y., Cai, T., Leach, P., Gu, Y. & Albright, S. (1999). *Simple service discovery protocol.* IETF Draft, draft-cai-ssdp-v1-03.txt.

Guttman, E., & Kempf, J. (1999). Automatic discovery of thin servers: SLP, Jini, and the SLP-Jini Bridge. In *Proceedings of the 25th Annual Conference of IEEE Industrial Electronics Society (IECON'99),* Piscataway, USA.

Guttman, E., Perkins, C., Veizades, J., & Day, M.(1999). *Service location protocol, version 2. IETF* (RFC 2608). Retrieved from http://www.ietf.org/rfc/rfc2608.txt

Hodes, T. D., Czerwinski, S. E., Zhao, B. Y., Joseph, A. D., & Katz, R. H. (2002, March/May). An architecture for secure wide-area service discovery. *ACM Wireless Networks Journal, 8*(2-3), 213-230.

Jetcheva, J., Hu, Y., Maltz, D., & Johnson, D. (2001, July). *A simple protocol for multicast and broadcast in mobile ad hoc networks.* Internet Draft draft-ietfmanet-simple-mbcast-01.txt, Internet Engineering Task Force.

Koubaa, H. (2003). *Localisation de services dans les réseaux ad hoc.* PhD thesis, Université Henri Poincaré Nancy,1, Mars 2003.

Koubaa, H., & Fleury, E. (2001, November). *A fully distributed mediator based service location protocol in ad hoc networks.* In IEEE Symposium on Ad hoc Wireless Networks, Globecom, San Antonio, TX.

Koubaa, H., & Fleury, E. (2002, July). *Service location protocol overhead in the random graph model for ad hoc networks.* In the IEEE Symposium on Computers and Communications, Taormina/Giardini Naxos, Italy.

Koubaa, H., & Wang, Z. (2004, June). *A hybrid content location approach between structured and unstructured topology.* In the 3rd Annual Mediterranean Ad hoc Networking Workshop, Bodrum, Turkey.

Lim, H., & Kim, C. (2000, August). *Multicast tree construction and flooding in wireless ad hoc networks.* In ACM MSWiM, Boston.

Nath, B., & Niculescu, D. (2003). Routing on a curve. *SIGCOMM Computer Communication Review, 33*(1), 155-160.

Peng, W., & Lu, X. (2000, August). *On the reduction of broadcast redundancy in mobile ad hoc networks.* In the 1st ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc), Boston.

Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., & Balakrishnan H. (2001). Chord: A scalable peer-to-peer lookup service for internet applica-

tions. In *Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (pp. 149-160). ACM Press.

Sun Microsystems Inc. (2003). *Jini technology core platform specification, version 2.0.* Retrieved June, 2003, from http://www.jini.org/nonav/standards/davis/doc/specs/html/core-title.html

Tchakarov, T., & Vaidya, N. (2004, January). *Efficient content location in wireless ad hoc networks.* In the IEEE International Conference on Mobile Data Management (MDM).

Tseng, Y., Ni, S., Chen, Y., & Sheu, J. (1999, August). The broadcast storm problem in a mobile ad hoc network. *5th Annual International Conference on Mobile Computing (MOBICOM),* Washington, DC, *31*(5), 78-91.

Wang, Z. (2003). *An agent-based integrated service platform for wireless and mobile environments.* Aachen, Germany: Shaker Verlag.

Wang, Z., & Seitz, J. (2002). An agent based service discovery architecture for mobile environments. In *Proceedings of the 1st Eurasian Conference on Advances in Information and Communication Technology,* Shiraz, Iran, October (LNCS 2510, pp. 350-357). Springer-Verlag.

Wang, Z., & Seitz, J. (2002, October). Mobile agents for discovering and accessing services in nomadic environments. In *Proceedings of the 4th International Workshop on Mobile Agents for Telecommunication Applications,* Barcelona, Spain (LNCS 2521, pp. 269-280). Springer-Verlag.

Williams, B., & Camp. (2002, June). *Comparison of broadcasting techniques for mobile ad hoc networks.* In the 3rd ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc), Lausanne, Switzerland.

Xu, D., Nahrstedt, D., & Wichadakul, D. (2000). *MeGaDiP: A wide-area media gateway discovery protocol.* In the 19th IEEE International Performance, Computing, and Communications Conference (IPCCC 2000).

Yang, X. W. (2001). A framework for semantic service discovery. In *Proceedings of the Student Oxygen Workshop, MIT Oxygen Alliance, MIT Computer Science and Artificial Intelligence Laboratory, 2001.* Retrieved from http://sow.csail.mit.edu/2001/proceedings/yxw.pdf

Zhao, W., & Guttman, E. (2000). *mSLP–Mesh enhanced service location protocol.* Internet Draft draft-zhao-slp-da-interaction-07.txt.

KEY TERMS

Aggregation: A process of grouping distinct data. Two different packets containing different data can be aggregated into a single packet holding the aggregated data.

Broadcast: A communication method that sends a packet to all other connected nodes on the network. With broadcast, data comes from one source and goes to all other connected sources at the same time.

Clustering: Identifying a subset of nodes within the network and vest them with the responsibility of being a cluster-head of certain nodes in their proximity.

Hash: Computing an address to look for an item by applying a mathematical function to a key for that item.

Mobile Ad Hoc Network: A kind of self-configuring mobile network connected by wireless links where stations or devices communicate directly and not via an access point. The nodes are free to move randomly and organize themselves arbitrarily, thus, the network's topology may change rapidly and unpredictably.

Multicast: A communication method that sends a packet to a specific group of hosts. With

multicast, a message is sent to multiple destinations simultaneously using the most efficient strategy that delivers the messages over each link of the network only once and only creates copies when the links to the destinations split.

Scalability: The ability to expand a computing solution to support large numbers of components without impacting performance.

Service: An abstraction function unit with clearly defined interfaces that performs a specific functionality. Users, applications, or other services can use the service functionality through well-known service interfaces without having to know how it is implemented.

Service Directory: An entity in service discovery architecture that collects and stores information about a set of services within a certain scope, which is used for searching and/or comparing services during the service discovery procedure. Service directory is also known as service repository or directory agent. Service directory can be organized in central or distributed manner.

Service Discovery: The activity to automatically find out servers in the network based on the given service type and service attributes. The service discovery is, therefore, a mapping from service type and attributes to the set of servers.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 165-178, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.12
Multimedia Messaging Peer

Kin Choong Yow
Nanyang Technological University, Singapore

Nitin Mittal
Nokia Pte Ltd, Singapore

ABSTRACT

The evolution in mobile messaging and mobile devices has made it possible to provide multimedia rich messaging capabilities to personal digital assistants (PDAs). The need for this arises simply because mobile service providers want to provide an enhanced messaging experience to the user. It also opens up new avenues for business, such as a shopping mall scenario. This chapter discusses the development of a multimedia messaging client for a PDA and a kiosk providing multimedia messages composition, search, share, and sending capabilities. This chapter also discusses the various messaging technologies, enabling wireless technologies, and the peer-to-peer model. The peer-to-peer technology used was Jxta, an XML-based and language agnostic peer-to-peer platform specification from Sun Microsystems. The peers (PDA client and the kiosk) were implemented using the application programming interfaces provided by the Personal Java reference implementation and the Jxta platform's Personal Java port.

INTRODUCTION

Over the years, mobile messaging has become an essential means of communication and is going to be even more so with the merging of the Internet and mobile networks. The ability to message from a phone to a computer on the Internet and vice versa is making messaging a powerful means of communication (Yeo, Hui, Soon, & Lau, 2001).

Mobile messaging is going through a gestation period (Tan, Hui, & Lau, 2001). Beginning with the enormously popular short message service (SMS) for simple text messages, the exciting enhanced message service (EMS) for illustrated text messages with sound entered the market as the second step. Multimedia messaging service (MMS) will become the ultimate messaging application, allowing users to create unique messages using various types of multimedia.

Initially, only mobile phones will support multimedia messaging. Personal digital assistants (PDA) are becoming very common and it is natural to think of putting this capability in them as well. PDAs have a larger form factor, memory footprint, and more powerful processors

and as a result would provide a richer messaging experience.

This chapter discusses the development of a multimedia messaging client for a PDA and a kiosk providing multimedia messages composition, search, share, and sending capabilities. Various messaging technologies, enabling wireless technologies, and the peer-to-peer model were also discussed and evaluated in this chapter. We substantiate the ideas discussed in this chapter with a description of the design and implementation of an MMS PDA client application with specific references to a shopping mall scenario.

BACKGROUND

Short Messaging Service

Text messaging uses the short messaging service (100-200 characters in length) and involves sending text messages between phones. Examples include "C U L8ER" and "OK. AT FLAT OR OFFICE." It is quick and dirty, hard to use the keypad, abrupt, punctuation challenged, and incredibly useful and popular. Text messaging also has a lot of advantages such as convenience, available on all phones, and discrete.

Text messaging is something that is most prevalent in the youth market and especially teenagers, who are able to manipulate the difficulty of entering text with the mobile phone keypad. In fact, it is suspected that this steep learning curve and the necessary insider knowledge are some of the things that appeal to the youngsters (Bennett & Weill, 1997).

SMS Advantages

Today's SMS has several advantages inherent in its fundamental features:

- **Store and forward.** This means that in the case that the recipient is not available, the short message is stored. Once the data is prepared and ready to send, SMS has advantages over packet data in that the burden of delivering the data is on the SMS center rather than the end user. The transaction costs incurred by the sender using SMS are therefore likely to be lower than a GPRS transaction.
- **Confirmation of delivery.** This means that the user knows that the short message has arrived. In the circuit switched data environment, there is an end-to-end connection and therefore the user knows that a connection has been established and the data is being transferred. In a GPRS environment, however, the concept is "always on" and this requires the user to find out whether the data has been sent or received.

SMS Disadvantages

However, today's SMS also has several disadvantages:

- **Limited message length.** The unit short message length is currently limited to 140 octets because of limitations in the signalling layer. It would be preferable to have a length that is several times this magnitude. Packet data services such as GPRS simplify non-voice transactions over mobile networks because the amount of data that can be communicated in any one session is significantly higher.
- **Inflexible message structure.** The structure of the SMS protocol data unit as defined in the GSM 03.40 standard is inflexible. The data coding scheme, origination address, protocol identifier, and other header fields are fixed—this has constrained the number of possible scenarios that can be indicated when developing applications.
- **Relatively slow signalling channel.** The latency turnaround time of services such

as GPRS and unstructured supplementary services data (USSD) tends to be faster than that for SMS. The signalling channel is used for several other purposes besides SMS, such as locating phones and managing call completion. Indeed, as SMS traffic volume grows, network operators have expressed some concern about potential service outages due to the overuse of and corresponding degradation in the scarce signalling resources.

- **Always store and forward.** Today's SMS is designed such that every short message always passes through the SMS center. Variations on this have been discussed at the UMTS committee level, such as forwarding messages and optionally storing them: immediately attempt delivery and if the message cannot be delivered, then store it. This reduces the processing power needed by the SMS center.

Enhanced Messaging Service

The enhanced messaging service (EMS) is the ability to send a combination of simple melodies, pictures, sounds, animations, modified text, and standard text as an integrated message for display on an EMS-compliant handset. For example, a simple black-and-white image could be displayed along with some text and a melody could be played at the same time. EMS is an enhancement to SMS but is very similar to SMS in terms of using the store-and-forward SMS centers and the signalling channel.

EMS Advantages

With the new, powerful EMS functionality, mobile phone users can add life to SMS text messaging in the form of pictures, animations, sound, and formatted text. As well as messaging, users will enjoy collecting and swapping pictures and ring signals and other melodies, downloading them from the Internet, or editing them directly on the phone.

- **Familiar user interface (UI).** Users will find EMS as easy to use as SMS. EMS provides a familiar user interface and compatibility with existing phones.
- **Compatible with SMS standards.** An EMS message can be sent to a mobile phone that does not support EMS or only partially supports EMS. All the EMS elements are located in the message header. A receiving phone that does not support the standard will ignore the EMS contents. Only the text message will be displayed to the receiver.
- **No new network infrastructure needed.** The beauty with EMS is that it uses existing SMS infrastructure and industry standards, keeping investments to a minimum for operators.
- **Concatenated messages.** SMS concatenation—stringing several short messages together—will be the key technical feature to enable the enhanced messaging service for the simple reason that complicated enhanced message designs, such as sending every alternate character in bold format, would occupy a large number of octets.

Multimedia Messaging Service

Overview of MMS

The multimedia messaging service, as its name suggests, is the ability to send and receive messages comprised of a combination of text, sounds, images, and video to MMS-capable handsets (MMS architecture, 2002). The trends for the growth in MMS are taking place at all levels within GSM (Patel & Gaffney, 1997), enabling technologies such as GPRS, EDGE, 3G, Bluetooth (http://www.bluetooth.com/), and wireless access protocol (WAP; WAP & MMS specifications, 2002).

MMS, according to the 3GPP standards (3GPP TS 23.140, 2002) is "a new service, which has no direct equivalent in the previous ETSI/GSM world or in the fixed network world." Here is an introduction to the features of this innovative new service:

- MMS is a service environment that allows different kinds of services to be offered, especially those that can exploit different media, multimedia, and multiple media.
- MMS will enable messages to be sent and received using lots of different media, including text, images, audio, and video.
- As more advanced media become available, more content-rich applications and services can be offered using the MMS service environment without any changes.
- The multimedia messaging service (MMS) introduces new messaging platforms to mobile networks in order to enable MMS. These platforms are the MMS relay, MMS server, MMS user databases, and new WAP gateways.
- MMS will require not only new network infrastructure but also new MMS-compliant terminals. MMS will not be compatible with old terminals, which means that before it can be widely used, MMS terminals must reach a certain penetration.

MMS Versus E-Mail

Four key areas best illustrate the differences between MMS and e-mail messages:

- **Message creation.** Conventional e-mail evolved from a simple text-based communication tool to include multimedia content, primarily in the form of attachments. A mobile device's form factor and the use of mobile data service, such as e-mail, become extremely unwieldy for the end user. MMS has been standardized specifically to take the limitations of mobile devices and the mobile end-users' needs into account.
- **Message delivery.** Conventional e-mail is designed around a store-and-retrieve model, whereas technologies such as SMS and MMS operate on a store-and-forward model. When sending a message using the former model, the sender must wait for the receiver to come online and access the network to retrieve the message. With the MMS model the message is stored and "pushed" or forwarded to the receiver immediately or as soon as the receiver comes online.
- **Messaging interoperability.** Today, e-mails can contain numerous and varied media formats and/or types. Each media type generally requires tools/plug-ins to render that media usable. In the desktop environment these tools are available locally or via the intra/Internet. This is not the case for the mobile subscriber due to the limited set of media-rendering capabilities available on mobile devices.
- **Message billing.** Conventional e-mail billing has relied on the subscription and/or access time model for billing the subscriber, resulting in little or no revenue for the service providers in some markets. SMS, however, has worked using a per-message billing model. Since MMS is the service evolution of SMS, mobile operators can apply SMS-like billing models to MMS without adverse customer reaction.

Implications of SMS on MMS

The current short message service has some unique advantages that other non-voice services do not have, such as store and forward and confirmation of message delivery. However, SMS also has some disadvantages, such as limited message length, inflexible message addressing structures, and signalling channel slowness.

Table 1. SMS versus MMS

Feature	**SMS**	**MMS**
Store and Forward (non-real-time)	Yes	Yes
Confirmation of message delivery	Yes	Yes
Communications type	P erson to person	Person to person
Media supported	T ext plus binary	Multiple— text, voice, video
Delivery mechanism	Signalling channel	Data traffic channel
Protocols	SMS specific, e.g., SMPP	General Internet, e.g., MIME, SMTP
Platforms	SMS center M	MS relay plus others
Applications	Simple person to person	Still images

Implications of EMS on MMS

Messaging will certainly develop beyond SMS. It is clear that an elegant solution like EMS that builds on simple text messaging and adds sound and simple images is a useful and powerful development that takes us beyond the limited reach of smart messaging services. It is clear that multimedia messaging service provides an ideal migration path to take advantage of the capacity and bandwidth that 3G/UMTS networks supply.

To summarise, MMS offers the content richness of e-mail and the instantaneous delivery of SMS messaging. While e-mail provides greater levels of utility, it fails to provide the instantaneous sharing and presentational capabilities needed to make it truly the mobile mass-market service. The strength of MMS in these areas make MMS the premium mass-market messaging service in 2.5G and 3G networks.

P2P MODEL AND JXTA

The previous section provides a discussion on the various messaging technologies, which are only the tools that MMS peers use to deliver their contents. What is more important is the distributed peer-to-peer model, which allows various MMS peers to discover and communicate with each other. This section gives a brief overview of distributed computing models and describes in detail one peer-to-peer technology, i.e., Jxta.

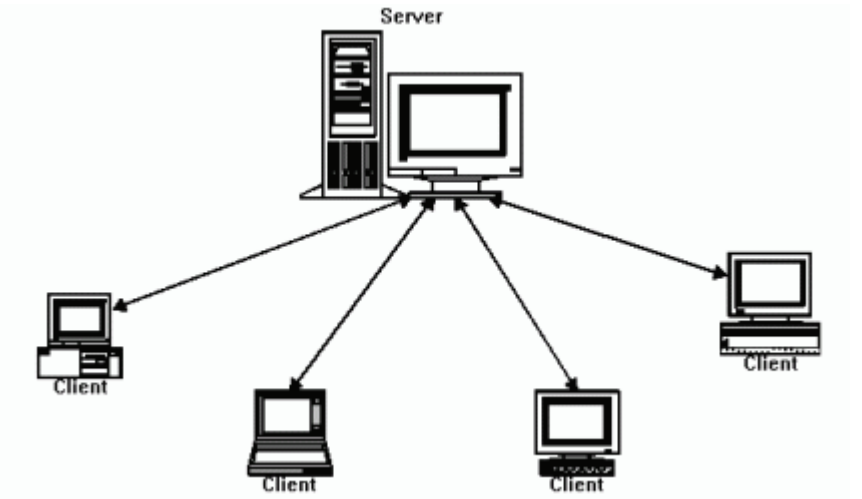
Distributed Computing Models

Overview of Client-Server Model

Today, the most common distributed computing model is the client-server model (Chambers, Duce, & Jones, 1984). Figure 1 depicts the typical client-server architecture.

In the client-server architecture, clients request services and servers provide those services. A variety of servers exist in today's Internet––Web servers, mail servers, FTP servers, etc. The client-server architecture is an example of a centralized architecture, where the whole network depends on central points to provide services. Regardless of the number of clients, the network can exist only if a server exists (Berson, 1992).

Figure 1. Client-server model

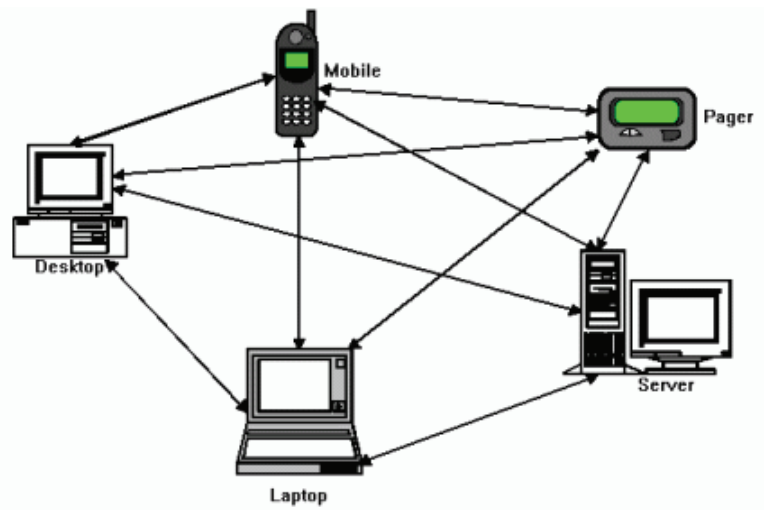


Overview of Peer-to-Peer Model

Like the client-server architecture, P2P is also a distributed computing model (Yemini, 1987). However, the P2P architecture is a decentralized architecture where neither client nor server status exists in a network (Madron, 1993). Every entity in the network, referred to as a *peer*, has equal status, meaning that an entity can either request a service (a client trait) or provide a service (a server trait). Figure 2 illustrates a P2P network.

Though peers all have equal status in the network, they don't all necessarily have equal physical capabilities. A P2P network might consist of peers with varying capabilities, from mobile devices to mainframes (Budiarto & Masahiko, 2002). A mobile peer might not be able to act as a server due to its intrinsic limitations, even though the network does not restrict it in any way.

Figure 2. P2P model



Comparison of Models

Both networking models feature advantages and disadvantages. One can visualize from Figure 1 that as a client-server network grows (i.e., as more and more clients are added), the pressure on the central point, the server, increases (giving rise to "hot spots"). As each client is added, the central point weakens and its failure can destroy the whole network.

A P2P network delivers a quite different scenario. Since every entity (or peer) in the network is an active participant, each peer contributes certain resources to the network, such as storage space and CPU cycles. As more and more peers join the network, the network's capability increases. A P2P network also differs from the client-server model in that the P2P network can be considered *alive* even if only one peer is active. The P2P network is unavailable only when no peers are active.

Jxta

Different protocols, different architectures, different implementations. That accurately describes current P2P solutions. Currently, developers use diverse methodologies and approaches to create P2P applications. Standards, abundant in the client-server world, are noticeably absent in the P2P world. To tackle this deficit, Sun developed *Jxta* (http://www.jxta.org/.).

Jxta strives to provide a base P2P infrastructure over which other P2P applications can be built (Project Jxta, 2002). This base consists of a set of protocols that are language independent, platform independent, and network agnostic. These protocols address the bare necessities for building generic P2P applications (Jxta technology overview, 2002). Designed to be simple with low overheads, the protocols target to build, to quote the Jxta vision statement, "every device with a digital heartbeat."

Jxta currently defines six protocols (Jxta protocol specification, 2002), but *not all Jxta peers are required to implement all six of them.* The number of protocols a peer implements depends on that peer's capabilities. A peer could use just one protocol. Peers can also extend or replace any protocol, depending on its particular requirements (Brookshier, 2002).

Jxta Core Building Blocks

Peer and Peer Groups. With today's limit on connectivity and available bandwidth, harnessing the entire Internet as one huge P2P network is impractical. Instead, some partitioning is

Figure 3. Jxta software architecture

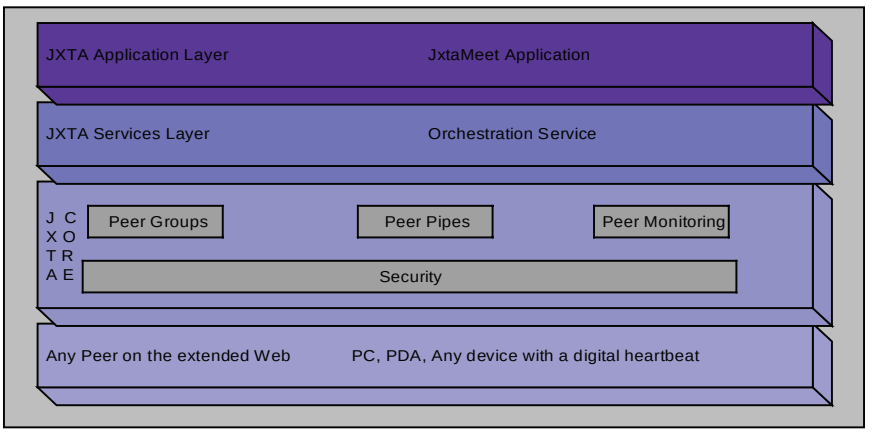


Table 2. Jxta services

Service name	Description
Pipe	The main means of communications between peers; provides an abstraction for a one-way, asynchronous conduit for information transfer.
Membership	Determines which peers belong to a peer group; handles arrival and departure of peers within a peer group.
Access	Security service for controlling access to services and resources within a peer group; a sort of security manager for the peer group.
Discovery	A way peers can discover each other, the existence of other peer groups, pipes, services, and the like.
Resolver	Allows peers to refer to each other, peer groups, pipes, or services indirectly through a reference (called an *advertisement* in Jxta lingo); the resolver binds the reference to an implementation at run time.

necessary. The logical partitioning of the physical network creates working sets of peers called *peer groups*. Peer group memberships can overlap with no restriction; in other words, any peer can belong to as many peer groups as necessary. The Jxta specification does not dictate or recommend an appropriate way of forming peer groups. In a Jxta network, a peer group is a collection of peers that share resources and services.

Jxta Services. Jxta *services* are available for shared use by peers within a peer group. In fact, a peer may join a group primarily to use the services available within that group. A set of services, called *core services,* is essential to the basic operation of a Jxta network. One instance of a core service already seen is the membership service. Table 2 shows the core services included in version 1.0 of the Jxta specification.

The initial reference implementation of project Jxta does not provide any services beyond the five listed in Table 2. Even some of the core services, such as the access service dealing with security, are implemented in a very basic way.

Pipes. One way to transfer data, files, information, code, or multimedia content between peers is through logical pipes, as defined by the Jxta specification. Jxta pipes are used to send messages (with arbitrary content) between peers. A pipe instance is, logically speaking, a resource within a peer group. The actual implementation of a pipe instance is typically through the pipe service. Unlike conventional UNIX-like systems, Jxta pipes are unidirectional and asynchronous.

Regardless of the type of pipe, the blocks of information carried though the pipe are referred to as Jxta *message*.

Messages. Jxta messages are data bundles that are passed from one peer to another through pipes. The Jxta specification is again as generic as possible here so as not to inadvertently introduce any implementation-dependent policies into the definition of a message. A message is defined to be an arbitrarily sized bundle, consisting of an envelope and a body. The envelope is in a standard format that contains:

- A header
- Source endpoint information (in URI form)
- Destination endpoint information (in URI form)
- A message digest (optional––for security purposes)

The body of a message is of arbitrary length and can contain an optional credential (for security purposes) and content.

Advertisements. Advertisements are the less obvious cousins of messages. Jxta advertisements are also XML documents. The content of an advertisement describes the properties of a Jxta component instance, such as a peer, a peer group, a pipe, or a service. *For example, a peer having access to an advertisement of another peer can try to connect directly to that other peer.* A peer having access to an advertisement of a peer group can use the advertisement to join that group. The current Internet analogue to an advertisement is the domain name and DNS record of a Web site.

JXTA Versus .NET and JINI

Jxta's XML-based messaging is similar to Microsoft's .Net and SOAP technologies. But that is a very thin foundation for comparison. As more and more third-party protocols make use of XML, it will become obvious that just using XML as a message format says nothing at all about an actual networking technology. Comparing the high-level, policy-rich, Web-services-based infrastructure that is .Net to the low-level, fundamental, policy-neutral nature of Jxta is a futile exercise.

Project Jxta and the Jini project are also fundamentally different. Both of them have some similarity in higher-level interaction, enabling true distributed computing over a network. However, the similarity abruptly ends there. Strategic differences between the two are: Jxta started life as a completely interoperable technology (any platform, any programming language, any vendor). Sun is a mere contributor in the community. Jini is a Java-centered technology that Sun will integrate and deploy strategically in future product offerings. Sun will maintain a degree of control over Jini's evolution.

MMS Kiosk and Jxta

In a P2P environment like Jxta, commonly accessed information gets replicated (the peers have a choice to keep a copy of content passing through them) and becomes available at peers fewer hops away. This avoids "hot spots" and is ideal for content sharing where the content can be of any type. For an MMS kiosk searching for multimedia messages, the situation is no different and it would thus be ideal to use a P2P framework to advertise and search for multimedia messages and media content.

A simple scenario is when a shop in a shopping mall adds more product brochures (in the form of MMS presentations) to its peer and advertises them. Subsequently, the kiosk would discover the new content and make it available to customers wishing to use the service from the kiosk. In a shopping mall this is likely to happen often as shops are always finding new ways of providing exciting offers to its customers. For example, shops providing multimedia messages for entertainment purposes would be updating their message collections often.

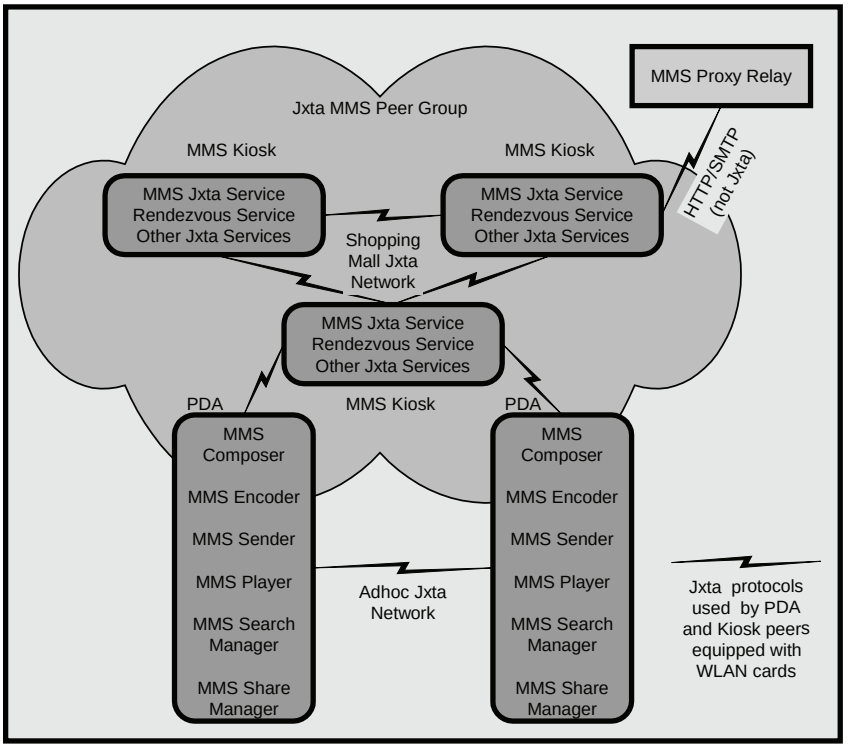
In another shopping mall scenario, a customer with a PDA walks into a shop, automatically connects to a *peer within range*, and begins searching for MMS messages and also an MMS service if it does not have the capability to connect to the MMS proxy-relay itself. In such a case, the PDA becomes a part of a peer group formed with the other customers around it. Hence, the search is not limited to the shopping mall peer-to-peer network but also to an ad hoc network created by the presence of other customers.

In a frequently changing environment like this, a P2P model like Jxta, with its dynamic resolution of peers and advertising and discovery capabilities, would be ideal to find the freshest content.

MMS Peer

Most shopping malls have information kiosks, which could be equipped with a network point and wireless service access capabilities using technologies like Bluetooth or 802.11b (IEEE 802.11, 2002). Location-specific MMS content

Figure 4. MMS peers and kiosk architecture



could be provided to customers in the vicinity of a kiosk or a shop. MMS would just be one of the many other services that can be provided at the kiosk, especially for mobile devices not connected to the cellular network or not subscribed to data services but with Bluetooth or 802.11b access.

The shops could provide product brochures, cards, postcards, pictures, comic strips, sounds, songs, etc. in the form of MMS presentations, which could be used by a customer to send to another person. The information kiosk could perform searches across the mall's network to update its multimedia message repository and provide a common contact point for all the shops in the mall.

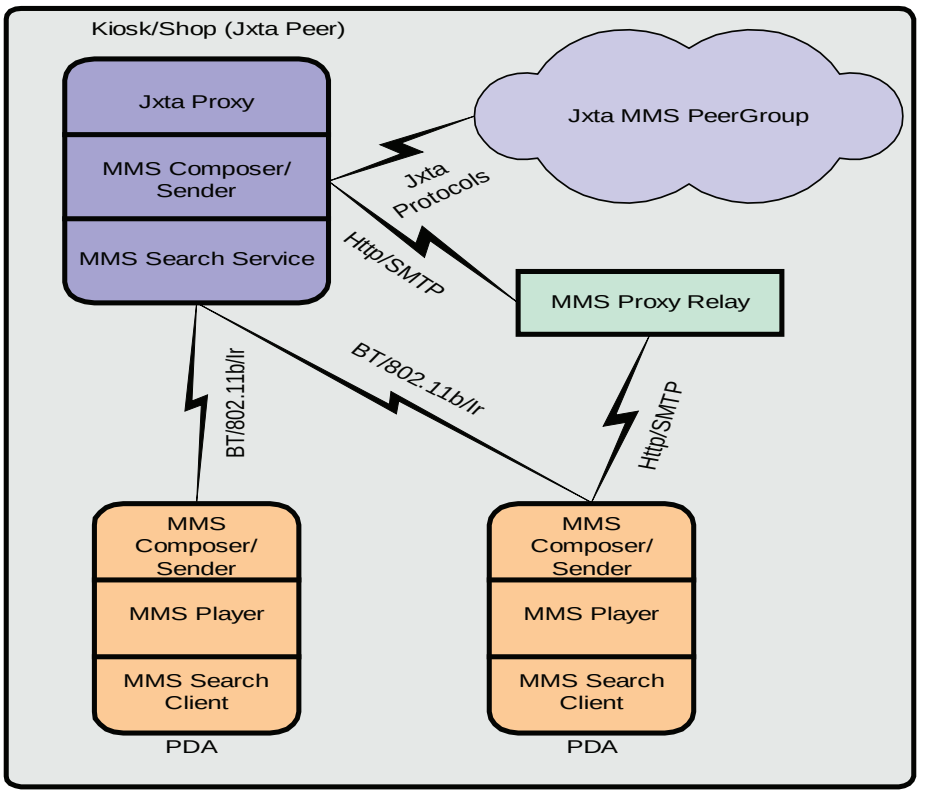
The requirement here is to provide enhanced customer service. The customer need not visit all the shops and need not verbally describe a product to another person before making a decision to buy something. Instead he or she could simply send the product brochures to the other person via MMS. These kiosks could also provide multimedia messages intended for fun and entertainment purposes and charge for them. Alternatively, shops could provide free MMS as a value-added service to their customers.

DESIGN AND IMPLEMENTATION OF MMS PEERS

Architecture I

Figure 5 shows the architecture of the whole MMS kiosk system when a PDA is incapable of running a Jxta peer. The PDA, in this architecture, is used to access the services provided by the kiosk/shop (like the MMS message search, composition, and sending). The PDA to kiosk/shop interaction is a client-server interaction. The kiosks and shops are the nodes in the P2P network here, not the customers' PDAs.

Figure 5. MMS kiosk environment architecture I



The PDA has been shown to have only an MMS composer, MMS player, and MMS sender. The MMS composer composes a message by aggregating all the media and presentation information provided by the user. The MMS sender performs a HTTP post to the MMS proxy-relay to send the message to its destination. An MMS player is also provided to the PDA client to view an MMS message before sending. The kiosk/shop is what provides the service to allow a customer to search for MMS messages and send them. The kiosk and the shops are part of a Jxta MMS peer group.

The protocols that the PDA can use to directly send to the MMS proxy are either HTTP or SMTP (if the MMS proxy-relay provides an SMTP interface). The communication between the kiosk/shop and the PDA can be over Bluetooth, IEEE 802.11b, or infrared. Infrared is not a good choice due to its very limited range (IrOBEX Specification, 2002).

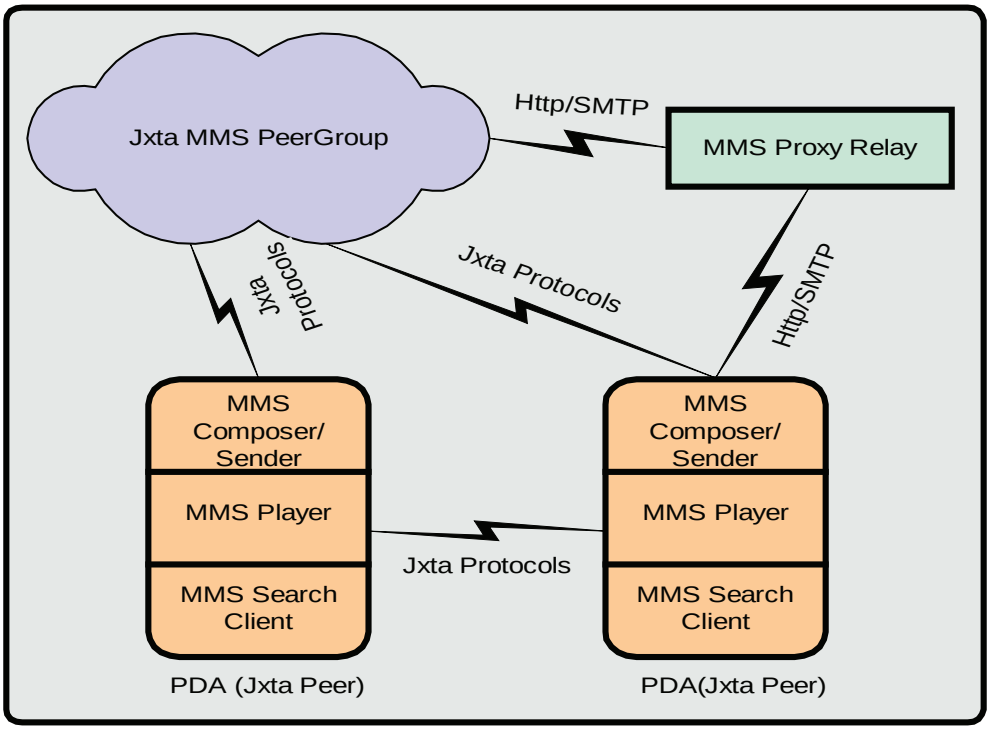
Architecture II

Figure 6 shows the final architecture due to the availability of the Jxta platform for a PDA (Jxta platform, 2002) and the P2P advantages it offers over Architecture I. Unlike Architecture I, this architecture is almost fully P2P, except the interaction between the Jxta peer and the MMS proxy-relay. This offers the advantage that a customer can become part of a peer group due to other customers around him. This opens opportunities for customers to exchange MMS messages they have on their devices.

PDA MMS Peer Design

The MMS peer on the PDA consists of four modules:

Figure 6. MMS kiosk environment architecture II



- MMS composer
- MMS encoder and sender
- MMS player
- MMS Jxta search and share

MMS Composer

This module allows a user to compose an MMS on the move (see Figure 7). It allows the user to select the media content and provide layout details and timing information for the slides of the MMS presentation. The process results in the generation of an SMIL file, which contains the presentation details of the media. Subsequently a Jar file (JAR documentation, 2002) is created with all the media files and the SMIL file. The MMS sender (in the next section) takes the Jar as its input, encodes it into an MMS, and sends it.

MMS Encoder and Sender

MMS can be sent either using HTTP Post or SMTP if the MMS proxy-relay provides both interfaces (see Figure 8). The two modes of sending the message could be chosen based on the priority of the message. Using SMTP takes longer to send, as the message has to be ultimately encoded according to MMS standards (MMS encapsulation specification, 2002). Hence, SMTP could be used to send low-priority messages.

MMS Player

The MMS player takes a Jar as input, extracts all the media and SMIL parts, and uses a SMIL parser to parse the SMIL and play it (see Figure 9). The slides in the SMIL presentation are rendered using *double buffering* (Double buffering in Java, 2002). The AMR audio is first converted to

Figure 7. MMS composer design

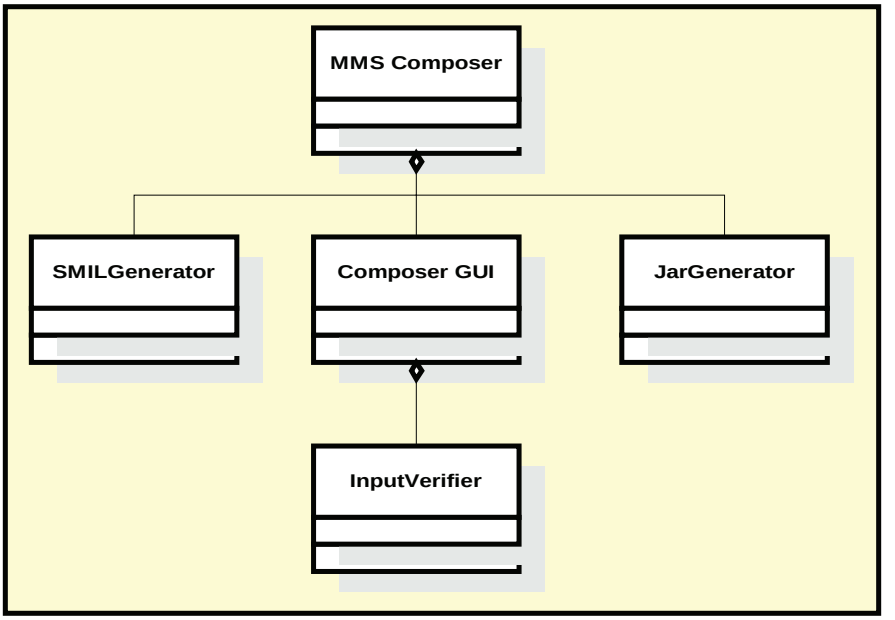
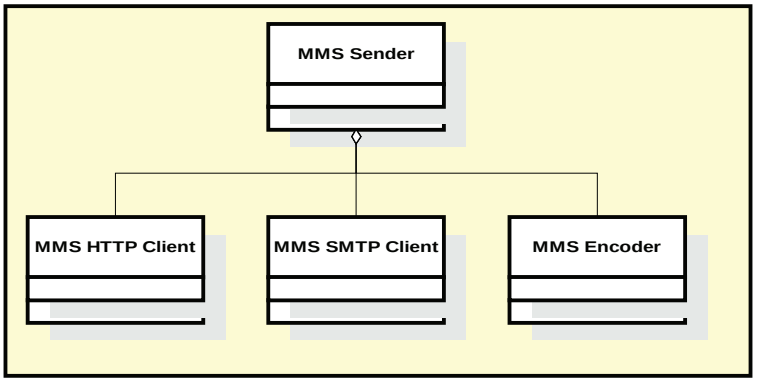


Figure 8. MMS sender design



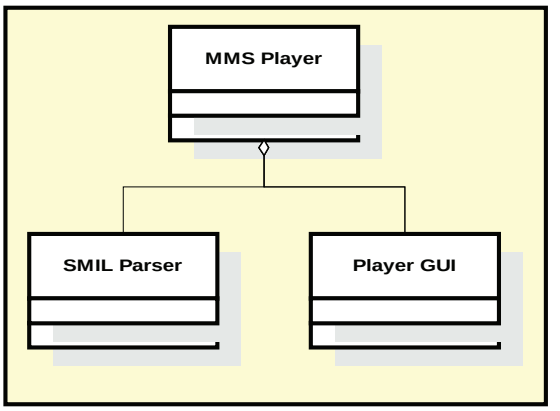
the WAV format and then played. Imelody (.imy files) files (iMelody specification, 2002) cannot be played by the application at this stage.

It was observed that some WAV files with a certain header were not playing on WinCE 3.0 but were playing on the desktop (Win 2000). The AMR to WAV converter native code was modified to take care of the change. The native code to convert and play is used via the Java Native Interface (JNI; Java Native Interface, 2002).

MMS Jxta Search and Share

The MMS application once started represents a Jxta peer. The peer becomes a member of a universal group called the net peer group. The peer then starts to discover its peers and available peer groups. The search and share module relies on two main modules called the peer group manager and the search manager (see Figure 10). The search and share uses the content management service.

Figure 9. MMS player design



A peer group manager is created after the start-up of the Jxta platform. This manager monitors and influences peer group discovery. It maintains the list of groups the peer has currently joined and allows one to join, leave, and create peer groups. The search manager is the interface to the file sharing application. The content put up locally by the peer or remotely in the peer group is described by an MMS advertisement. The MMS advertisement allows propagation along normal Jxta discovery channels.

MMS Kiosk Design

According to Architecture I, the kiosk consists of three modules: one to compose and send an MMS message, a second to search for MMS content, and a third to handle requests from the PDA. In Architecture II, the kiosk is not much different from the PDA peer except that it would also be expected to act as a rendezvous and run a MMS Jxta service described below.

MMS Jxta Service

Peers having access to the MMS proxy-relay provide the MMS Jxta service. The service is searched for by peers which do not have connection to the Internet and cannot access the MMS proxy-relay directly. The service advertisement contains the pipe advertisement so that other peers can connect to it and use it. This service should be run as a peer group service so that there is a higher chance of the availability of the service. When run as peer group service, the peers running the service rely on one peer group service advertisement instead of publishing their own for the same service.

Figure 10. MMS search and share design

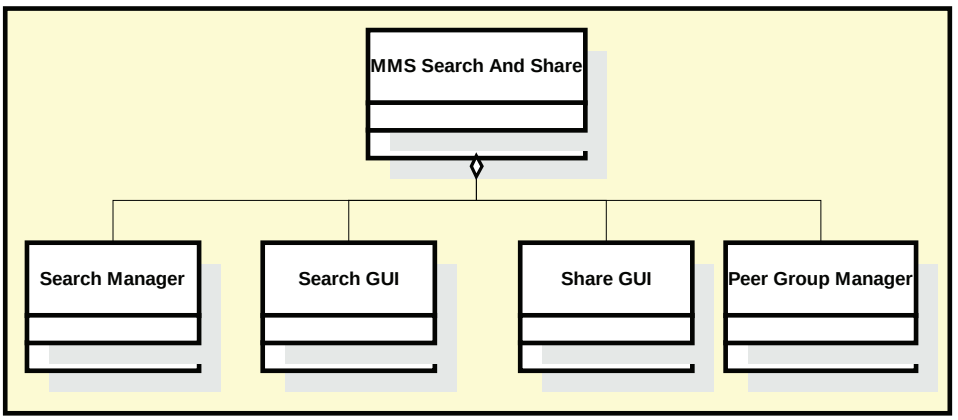


Figure 11. MMS Jxta service design

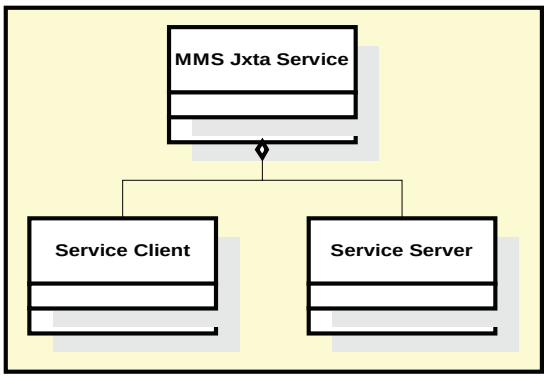
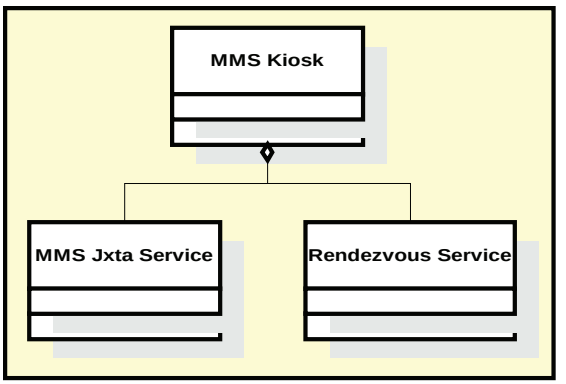


Figure 12. MMS kiosk design for Jxta-enabled PDAs



Jxta Proxy

Jxta proxy is required for devices that cannot run the Jxta core services yet. The Jxta proxy should receive the requests from the PDA client and then use the MMS search and messaging services on behalf of the PDA. It would also provide the interface to the transport protocol used between the kiosk/shop and the PDA.

Graphical User Interface Design

The graphical user interface (GUI) was designed keeping the PDA in mind. The user interface uses as many components that can be either easily clicked or tapped with a stylus. The following things were taken into consideration for the user interface design:

- A user would always want to have the list of peers and peer groups in front of him or her because of constant interaction with these entities.
- The limited screen size of the PDA requires that every function be provided without cluttering the screen. Thus every function is provided on a new screen.

Figure 13. Kiosk design for non-Jxta-enabled PDAs

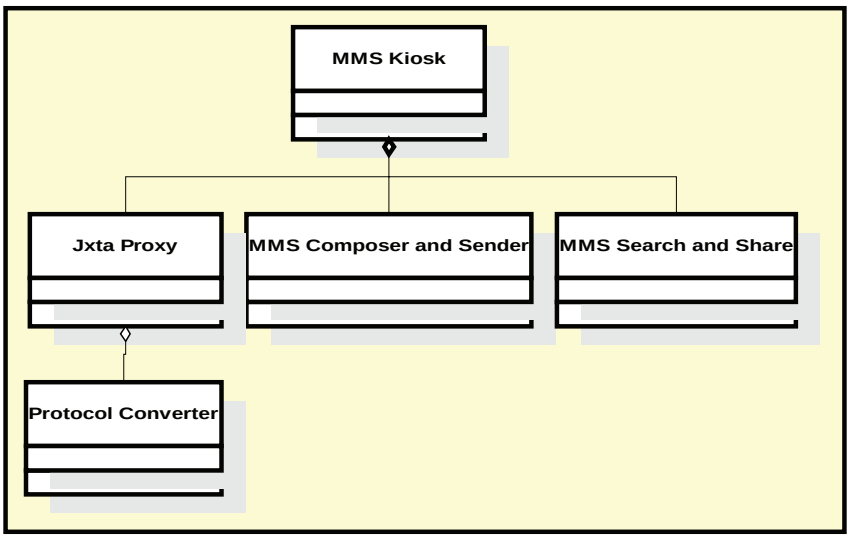


Figure 14. MMS player and sender

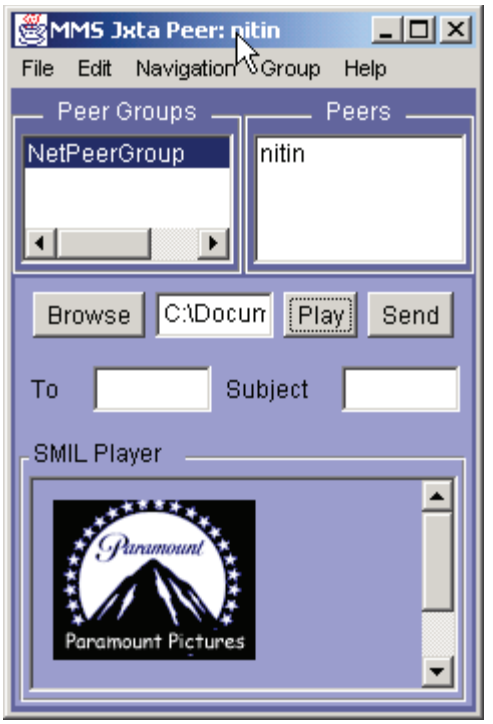
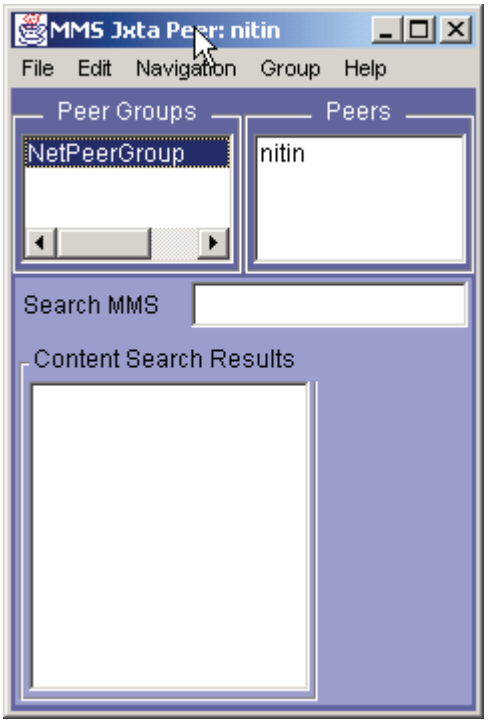


Figure 15. MMS search



- This sort of layout would be ideal if service clients are to be loaded dynamically upon discovery of a service.

Figure 14 shows the MMS player and sender GUI. It also shows the lists of peer groups and peers currently visible. The peer groups and peers lists keep getting updated automatically. Figure 15 shows the MMS search GUI. A user can enter the keywords and press enter to search. A button will be added also to allow easy use on the PDA.

Figure 16 shows the MMS share GUI. It is similar to the MMS search GUI. It allows search for content in the current peer group. Figure 17 shows the MMS composer GUI. It is used to compose an MMS using various media files input by the user slide by slide. Figure 18 shows the Navigation menu, which is used to navigate to the various GUI components seen earlier. The design is such that the same region is used for the various functions. Figure 19 shows the Group menu that

Figure 16. MMS share

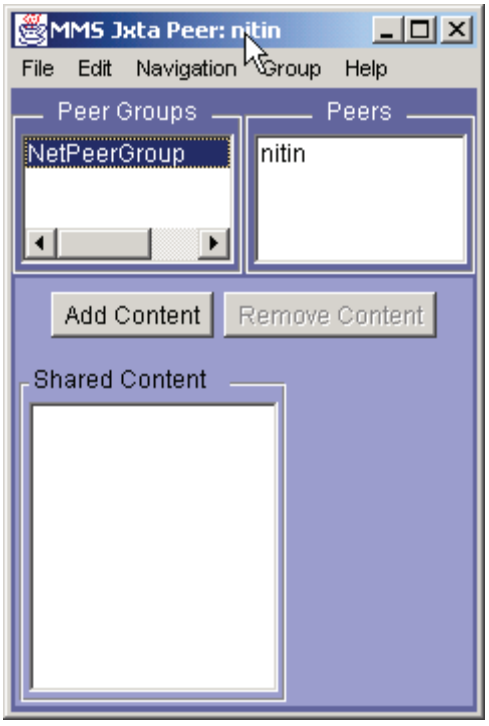


Figure 17. MMS composer

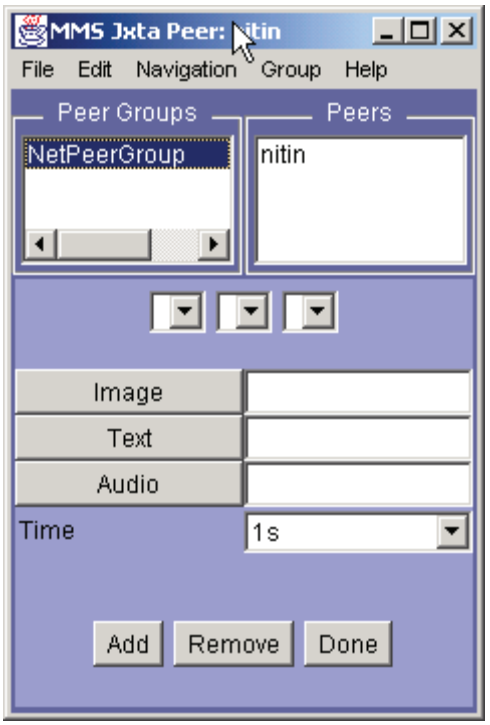
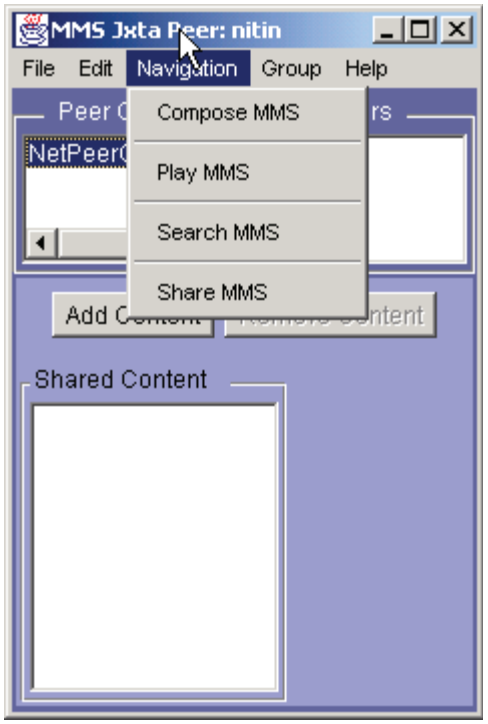


Figure 18. Navigation menu



allows one to create, join, and leave a peer group after selecting in the peer group list below.

Comparison with Other MMS Solutions

There are some other MMS clients for the PDA that exist now. The one from Anny Way (MMS—Opportunities, migration and profits, 2003) is specifically for Pocket PC. EasyMessenger (EasyMessenger, 2003) from Oksijen Technologies is the only other Personal-Java-based MMS client but without the additional P2P features provided by us. Electric Pocket's Pixer (Pixer MMS, 2003) is another MMS client that only supports Pocket PC and Palm OS and can be used to send images only. There seems to be no work done on using MMS and Jxta together or for that matter not even EMS or SMS and Jxta.

All the solutions above are MMS clients with a view to sending multimedia messages, a progress from SMS or EMS. The MMS peer was developed with a view to making not only messaging a more pleasant and easier experience but also to provide features that would facilitate access to a variety of content. The searching and sharing of content from peers in the vicinity (shoppers) as well as content stores (kiosks) make it a compelling multimedia messaging solution.

FUTURE TRENDS

MMS Message Receiver Module for the PDA

The next step is to implement a receiver module for the PDA so that the MMS peer is able to achieve two-way messaging. In the following discussion we give a flowchart to decode the message if retrieved directly from the MMS proxy-relay using the HTTP GET method. This flowchart will be helpful for the implementation of the module to receive MMS messages on a PDA.

MMS Main Header Decoder

Figure 20 is a flowchart of the decoding process for the main header of the "m-retrieve-conf" (MMS client transaction specification, 2002) message of content type *application/vnd.wap.multipart.related*. It assumes that after the sending of the HTTP GET request, the MMS proxy-relay will return with a message along with the HTTP headers. The HTTP headers can be easily skipped by looking for two consecutive carriage returns and line feed pairs. After this the encoded MMS header starts which are read byte by byte till the byte of number of body parts is reached.

The "Read uintvar length" process needs some further explanation. Note that this value can be variable in length and uses the *variable length unsigned integer encoding* as discussed in the design and implementation section. A byte follows if the byte currently read has its most significant bit equal to 1. So one knows how many bytes to read without actually knowing the length of this value. Also note that the byte denoting number of body parts used after the main headers is going to be deprecated and might not be the perfect way to reach an end. This should not affect the decoding though, as the last decision box should be reached only after all the headers, their parameters, and their values have been read. This flowchart assumes no user-defined headers.

The To and Subject header values can have length bytes preceding the value. This length value could either be 1 byte or encoded using *variable length unsigned integer encoding.* Its length in the latter case is followed by the charset for the following text.

MMS Body Header Decoder

In Figure 21 "hLen" is the length of the whole header and "cLen" is the length of the Content Type header. Note that the Content Type header in the body part is just the value of this header and is not preceded by a code for the header name.

Service Client Plug-In Feature

The service client plug-in feature refers to the client download option. The current implementation assumes the client for a service to be there on the peer. As the peer already has core Jxta functionalities, it is a good idea to use them to provide this feature. The advertisements of a service could specify the location of a client, which could be transferred over to the peer and dynamically loaded. This is possible in Java using the API for loading classes. To enable this feature one could create a Jxta service that has clients registered with it.

PDA to PDA Messaging

With the existing application framework, PDA to PDA MMS messaging can be easily enabled using the Jxta messaging layer. As PDAs are more capable than mobile phones, even video could be enabled for PDA to PDA messaging. All it would mean is using another media type in the SMIL or the encoded message.

To account for different PDAs communicating, the user agent profile specification (UAProf; WAP UAProf, 2002) could be used for capability negotiation. The UAProf schema for MMS characteristics (client transactions) could be adapted to the PDA situation. The XML messaging layer for Jxta would enable the use of this XML scheme effectively.

CONCLUSION

The Jxta platform Personal Java port came out very recently and the application was designed and implemented with it in mind. If the basic platform functionalities have been ported correctly, then it should not take long to port this whole application to the PDA. The application conforms to the Personal Java standard when checked with the compliance tool. This implies it should work on the PDA without requiring any changes.

Figure 20. MMS main header decoder flowchart

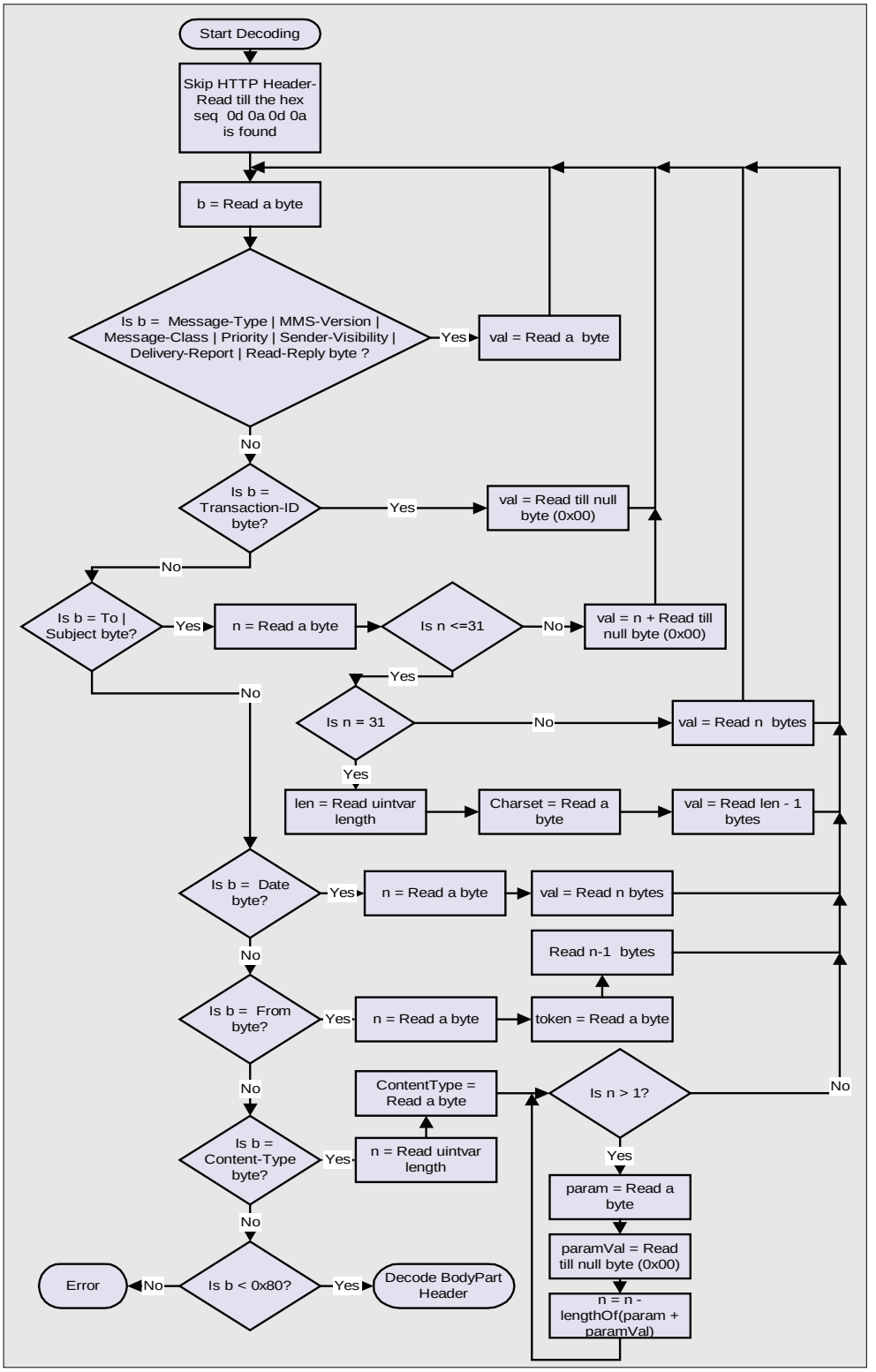
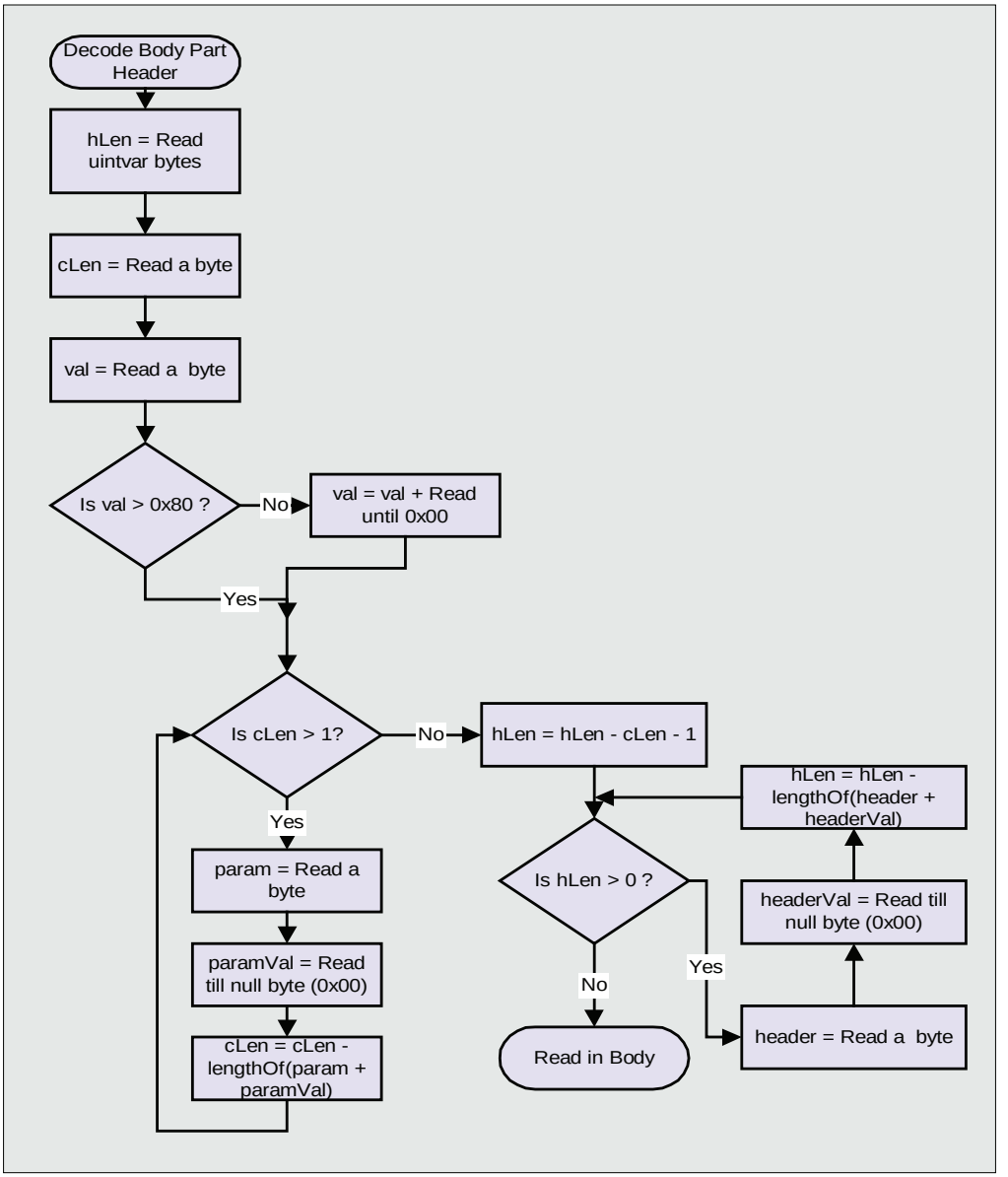


Figure 21. MMS body part header decoder flowchart



Currently the MMS client by itself works on the PDA. An MMS player and sharing of content were developed. The former was implemented while searching for a reasonably priced Bluetooth SDK for WinCE and trying various Bluetooth PCMCIA cards with the freely available Bluetooth stack called CStack (http://www.cstack.com).

The project will have a commercial value when shopping malls in Singapore install wireless networks or have wireless kiosks. This is the main driving force behind this project. Another driving force was to be able to pave a path to introduce multimedia messaging to students in Nanyang Technological University.

In conclusion, with the increase in memory and processing power of a plethora of mobile devices found in the market and the ongoing improvements in available bandwidth to the user, MMS

is a service to look forward to and more so with peer-to-peer technologies like Jxta, which will make it truly ubiquitous.

REFERENCES

3GPP TS 23.140: MMS functional description. (2002). Retrieved from http://www.3gpp.org/

Bennett, M., & Weill, P. (1997). Exploring the use of electronic messaging infrastructure: The case of a telecommunications firm. *Journal of Strategic Information Systems, 6*(1), 7-34.

Berson, A. (1992). *Client/server architecture*. New York: McGraw-Hill.

Brookshier, D. (2002). *JXTA: Java P2P programming.* Indianapolis, IN: Sams.

Budiarto, S. N., & Masahiko, T. (2002). Data management issues in mobile and peer-to-peer environments. *Data & Knowledge Engineering, 41*(2-3), 183-204.

Chambers, F. B., Duce, D. A., & Jones, G. P. (Eds.). (1984). *Distributed computing.* London; Orlando, FL: Academic Press.

Double buffering in Java. (2002). Retrieved from http://java.sun.com/docs/books/tutorial/extra/fullscreen/doublebuf.html

EasyMessenger. (2003). Retrieved from http://www.o2.com.tr/easymessenger.htm

ETSI GTS GSM 03.40. (2002). Retrieved from http://www.etsi.org/

IEEE 802.11. (2002). Retrieved from http://grouper.ieee.org/groups/802/11/index.html

IMelody specification. (2002). Retrieved from http://www.irda.org/standards/specifications.asp

IrOBEX specification. (2002). Retrieved from http://www.irda.org/standards/specifications.asp

JAR documentation. (2002). Retrieved from http://java.sun.com/products/jdk/1.1/docs/guide/jar/

Java Native Interface. (2002). Retrieved from http://java.sun.com/j2se/1.3/docs/guide/jni/

Jxta platform. (2002). Retrieved from http://platform.jxta.org

Jxta protocol specification v1.0. (2002). Retrieved from http://spec.jxta.org/v1.0/docbook/JXTAProtocols.html

Jxta technology overview. (2002). Retrieved from http://www.jxta.org/project/www/docs/TechOverview.pdf

Madron, T. W. (1993). *Peer-to-peer LANs: Networking two to ten PCs.* New York: Wiley.

MMS—Opportunities, migration and profits. (2003). Retrieved from http://www.annyway.com/annyway-com2.htm

Multimedia messaging service (MMS) architecture overview. (2002). Retrieved from http://www.wapforum.org/what/technical.htm

Multimedia messaging service (MMS) client transaction specification. (2002). Retrieved from http://www.wapforum.org/what/technical.htm

Multimedia messaging service (MMS) encapsulation specification. (2002). Retrieved from http://www.wapforum.org/what/technical.htm

Patel, A., & Gaffney, K. (1997). A technique for multi-network access to multimedia messages. *Computer Communications, 20*(5), 324-337.

Pixer MMS. (2003). Retrieved from http://electricpocket.com/products/carriers.html

Project Jxta: Getting started. (2002). Retrieved from http://www.jxta.org/project/www/docs/GettingStarted.pdf

Tan, D. H. M., Hui, S. C., & Lau, C. T. (2001). Wireless messaging services for mobile users. *Journal of Network and Computer Applications, 24*(2), 151-166.

WAP & MMS specifications. (2002). Retrieved from http://www.wapforum.org/what/technical.htm

WAP UAProf. (2002). Retrieved from http://www.wapforum.org

Yemini, Y. (Ed.). (1987). *Current advances in distributed computing and communications.* Rockville, MD: Computer Science Press.

Yeo, C. K., Hui, S. C., Soon, I. Y., & Lau, C. T. (2001). A unified messaging system on the Internet. *Microprocessors and Microsystems, 24*(10) 523-530.

This work was previously published in Mobile Commerce Applications, edited by N. S. Shi, pp. 203-230, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.13
Situated Multimedia for Mobile Communications

Jonna Häkkilä
Nokia Multimedia, Finland

Jani Mäntyjärvi
VTT Technical Centre of Finland, Finland

ABSTRACT

This chapter examines the integration of multimedia, mobile communication technology, and context-awareness for situated mobile multimedia. Situated mobile multimedia has been enabled by technological developments in recent years, including mobile phone integrated cameras, audio-video players, and multimedia editing tools, as well as improved sensing technologies and data transfer formats. It has potential for enhanced efficiency of the device usage, new applications, and mobile services related to creation, sharing, and storing of information. We introduce the background and the current status of the technology for the key elements constructing the situated mobile multimedia, and identify the existing development trends. Then, the future directions are examined by looking at the roadmaps and visions framed in the field.

INTRODUCTION

The rapid expansion of mobile phone usage during last decade has introduced **mobile communication** as an everyday concept in our lives. Conventionally, **mobile terminals** have been used primarily for calling and employing the short message service (SMS), the so-called text messaging. During recent years, the multimedia messaging service (MMS) has been introduced to a wide audience, and more and more mobile terminals have an integrated camera capable of still, and often also video recording. In addition to imaging functions, audio features have been added and many mobile terminals now employ (e.g., an audio recorder and an MP3 player). Thus, the capabilities of creating, sharing, and consuming multimedia items are growing, both in the sense of integrating more advanced technology and reaching ever-increasing user groups. The

introduction of third generation networks, starting from Japan in October 2001 (Tachikawa, 2003), has put more emphasis on developing services requiring faster data transfer, such as streaming audio-video content, and it can be anticipated that the role of multimedia will grow stronger in mobile communications.

The mobile communications technology integrating the **multimedia** capabilities is thus expected to expand, and with this trend both the demand and supply of more specific features and characteristics will follow. In this chapter we concentrate on describing a specific phenomenon under the topic of mobile multimedia—namely, integrating **context awareness** into mobile multimedia.

Context-awareness implies that the device is to some extent aware of the characteristics of the concurrent usage situation. Contextual information sources can be, for instance, characteristics of the physical environment, such as temperature or noise level, user's goals and tasks, or the surrounding infrastructure. This information can be bound to the use of mobile multimedia to expand its possibilities and to enhance the **human computer interaction**. Features enhancing context awareness include such things as the use of context-triggered device actions, delivery of multimedia-based services, exploiting recorded metadata, and so on.

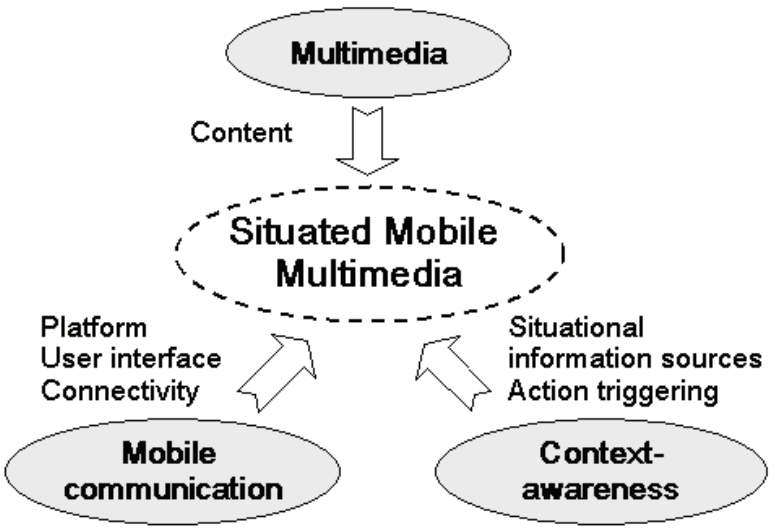
In this chapter we will look into three key aspects—mobile communications, multimedia, and context-awareness—and consider how they can be integrated. We will first look at each key element to understand the background and its current status, including identifying the current development trends. Then the future directions will be examined by looking at the roadmaps and visions framed in the field. The challenges and possibilities will then be summarized.

BACKGROUND

The development of digital multimedia has emerged in all segments of our everyday life. The home domain is typically equipped with affiliated gadgets, including digital TV, DVD, home theaters, and other popular infotainment systems. The content of the digital multimedia varies from entertainment to documentation and educative material, and to users' self-created documents. Learning tools exploiting digital multimedia are evident from kindergarten to universities, including all fields of education (e.g., language learning, mathematics, and history). Digital multimedia tools are used in health care or security monitoring systems. So far, the platforms and environments for the use of digital multimedia have been non-mobile, even "furniture type" systems (i.e., PC-centered or built around a home entertainment unit). The PC, together with the Internet, has been the key element for storing, editing, and sharing multimedia content. User-created documents have involved gadgets such as video cameras or digital cameras, from where the data needs to be transferred to other equipment to enable monitoring or editing of the produced material.

However, the role of mobile multimedia is becoming increasingly important both in the sense of creating, and sharing and monitoring the content. The increased flexibility of use following from the characteristics of a mobile communication device—it is a mobile, personal, and small-size gadget always with the user—has expanded the usage situations and created possibilities for new applications; the connection to the communication infrastructure enables effective data delivery and sharing. Adding the situational aspect to mobile multimedia can be utilized using context awareness, which brings information of the current usage situation or preferred functions, and can be used for action triggering for instance (Figure 1). Here, the multimedia content is dealt with in

Figure 1. Integrating multimedia, mobile communication technology, and context awareness



a mobile communication device, most often a mobile phone, which offers the physical platform and user interface (UI) for storing, editing, and observing the content.

In the following sections, we first look at the use of mobile communication technology and then the concept of context awareness more closely.

Mobile Phone Culture

During recent decades, mobile communication has grown rapidly to cover every consumer sector so that the penetration rates approach and even exceed 100% in many countries. Although the early mobile phone models were designed for the most basic function, calling, they soon started to host more features, first related primarily to interpersonal communication, such as phone-book and messaging applications accompanied by features familiar to many users from the PC world (e.g., electronic calendar applications and text document creation and editing). These were followed by multimedia-related features, such as integrated cameras, FM radios, and MP3 players, and applications for creating, storing, and sharing multimedia items.

Defining different audio alert profiles and ringing tones, and defining distinct settings for different people or caller groups, expanded the user's control over the device. Personalization of the mobile phone became a strong trend supported by changeable covers, operator logos, and display wallpapers. All this emphasized the mobile phone as a personal device.

The text messaging culture was quickly adopted by mobile phone users, especially by teenagers. The asynchronous communication enabled new styles of interaction and changed the way of communicating. Text messaging has been investigated both from the viewpoint of how it is associated with everyday life situations, the content and character of messaging, and the expression and style employed in messaging (Grinter & Eldridge, 2003). When looking at teenagers' messaging behavior, Grinter and Eldridge (2003) report on three types of messaging categories: chatting, planning activities, and coordinating communications in which the messaging leads to the use of some other communication media, such as face-to-face meetings or phone calls. Text messaging has also created its own forms of expression (e.g., the use of shortened words,

mixing letters, and number characters, and the use of acronyms that are understood by other heavy SMS users or a certain group).

Due to the novelty of the topic, mobile communication exploiting multimedia content has so far only been slightly researched. Cole and Stanton (2003) report that mobile technology capable of pictorial information exchange has been found to hold potential for a youngster's collaboration during activities, for instance, in story telling and adventure gaming. Kurvinen (2003) reports a case study on group communication, where users interact with MMS and picture exchange for sending congratulations and humorous teasing. Multimedia messaging has been used, for example, as a learning tool within a university student mentoring program, where the mentors and mentees were provided with camera phones and could share pictorial information on each other's activities during the mobile mentoring period (Häkkilä, Beekhuyzen, & von Hellens, 2004). In a study on camera phone use, Kindberg, Spasojevic, Fleck, and Sellen (2005) report on user behavior in capturing and sharing the images, describing the affective and functional reasons when capturing the photos, and that by far the majority of the photos stored in mobile phones were taken by the user him or herself and kept for a sentimental reason.

Context Awareness for Mobile Devices

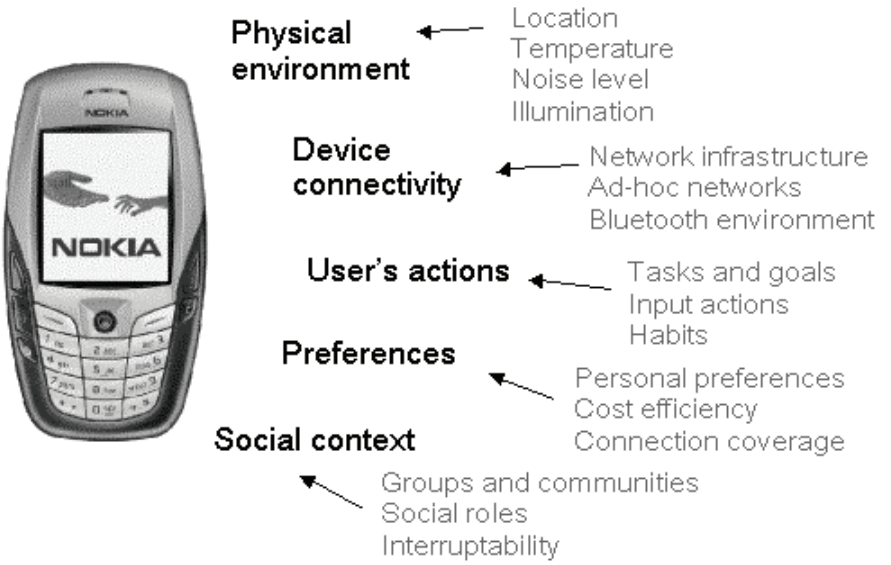
In short, context awareness aims at using the information of the usage context for better adapting the behavior of the device to the situation. Mobile handsets have limited input and output functionalities, and, due mobility, they are used in altering and dynamically varying environments. Mobile phones are treated as personal devices, and thus have the potential to learn and adapt to the user's habits and preferences. Taking these special characteristics of mobile handheld devices into account, they form a very suitable platform for context-aware application development. Context awareness has been proposed as a potential step in future technology development, as it offers the possibilities of smart environments, adaptive UI's, and more flexible use of devices.

When compared with the early mobile phone models of the 1990s, the complexity of the device has increased dramatically. The current models have a multiple number of applications, which, typically, must still be operated with the same number of input keys and almost the same size display. The navigation paths and the number of input events have grown, and completing many actions takes a relatively long time, as it typically requires numerous key presses. One motivation for integrating context awareness into mobile terminals is to offer shortcuts to the applications needed in a certain situation, or automating the execution of appropriate actions.

The research so far has proposed several classifications for contextual information sources. For example, the TEA (Technology for Enabling Awareness) project used two general categories for structuring the concept of context: human factors and physical environment. Each of these has three subcategories: human factors divides into information on the user, his or her social environment, and tasks, and physical environment distinguishes location, infrastructure, and physical conditions. In addition, orthogonal to these categories, history provides the information on the changes of context attributes over time Schmidt, Beigl, and Gellersen (1999). Schilit, Adams, and Want (1994) propose three general categories: user context, physical context, and computing context. Dey and Abowd (2000) define the context as "any information that can be used to characterize the situation of an entity."

In Figure 2, we present the contextual information sources as they appear from the mobile communication viewpoint, separating the five different main categories—physical environment, device connectivity, user's actions, preferences, and social context—which emphasize the spe-

Figure 2. Context-aware mobile device information source categories and examples



cial characteristics of the field. These categories were selected as we saw them to represent the different aspects especially important to the mobile communication domain, and they are briefly explained in the following. The proposed categories overlap somewhat and individual issues often have an effect on several others. Thus, they are not meant as strictly separate matters but aim to construct an entity of overall contextual information sources.

Physical environment is probably the most used contextual information source, where the data can be derived from sensor-based measurements. Typical sensor data used in context-aware research includes temperature, noise, and light intensity sensors and accelerometers. Location, a single attribute that has resulted in the most research and applications in the field of mobile context-aware research, can be determined with GPS or by using the cell-ID information of the mobile phone network. *Device connectivity* refers to the information that can be retrieved via data transfer channels that connect the device to the outside world, other devices, or the network infrastructure. This means not only the mobile phone network, such as GSM, TDMA, or GPRS connections, but also ad hoc type networks and local connectivity systems, such as Bluetooth environment or data transfer over infrared. A certain connectivity channel may enable different types of context-aware applications: for example, Bluetooth can be used as a presence information source due its short range. The category *user's actions* implies the user's behavior, which here covers a wide range of different aspects from single input events, such as key presses and navigation in the menus, to prevailing tasks and goals, and general habits typical of the individual user.

Contrary to the previous categories, which are more or less typical of the research in the field of context awareness, we propose that the last two categories have an important role when using a mobile communication device. By *preferences* and *social context*, we refer to the factors relating to the use situations especially important from the end-user perspective. The preferences cover such issues as cost-efficiency, data connection speed, and reliability, which are important to the end-user and which relate closely to the connectivity issues dealing with handovers and alternative data

transfer mediums. But, not only technical issues affect the usage. The user's personal preferences, which can offer useful information for profiling or personalizing mobile services, are also important. *Social context* forms an important information category as mobile terminals are still primarily used for personal communication and are often used in situations where the presence of other people cannot be avoided. This category forms a somewhat special case among the five classes (Figure 2) as it has a strong role both as an input and an output element of a context-aware application, so it can—and should—be taken into account both as an information source and in the consequent behavior of a context-aware system. By inferring the prevailing social context, one can not only gain information on the preferred functions but also vice versa–the social context also has an effect on how we wish the device to react in terms of interruptability or privacy.

Contextual information can be used for automating certain device actions: when specified conditions are fulfilled, the detected context information triggers a pre-defined action to take place. As an example, a mobile phone ringing tone volume could be automatically set to maximum if the surrounding noise exceeded 90 dB. In addition to automated device behavior, semi-automated and manual execution of actions has been suggested to ensure an appropriate level of user control (Mäntyjärvi, Tuomela, Känsälä, & Häkkilä, 2003). Previous work in the field of mobile context-aware devices has implemented location-aware tour guides and reminders (Davies, Cheverst, Mitchell, & Efrat, 2001), where accessing a certain location triggers the related information or reminder alert to appear on the device screen, or the automated ringing tone profile changes (Mantyjärvi & Seppänen, 2003), as does screen layout adaptation (Mäntyjärvi & Seppänen, 2003) in certain environmental circumstances.

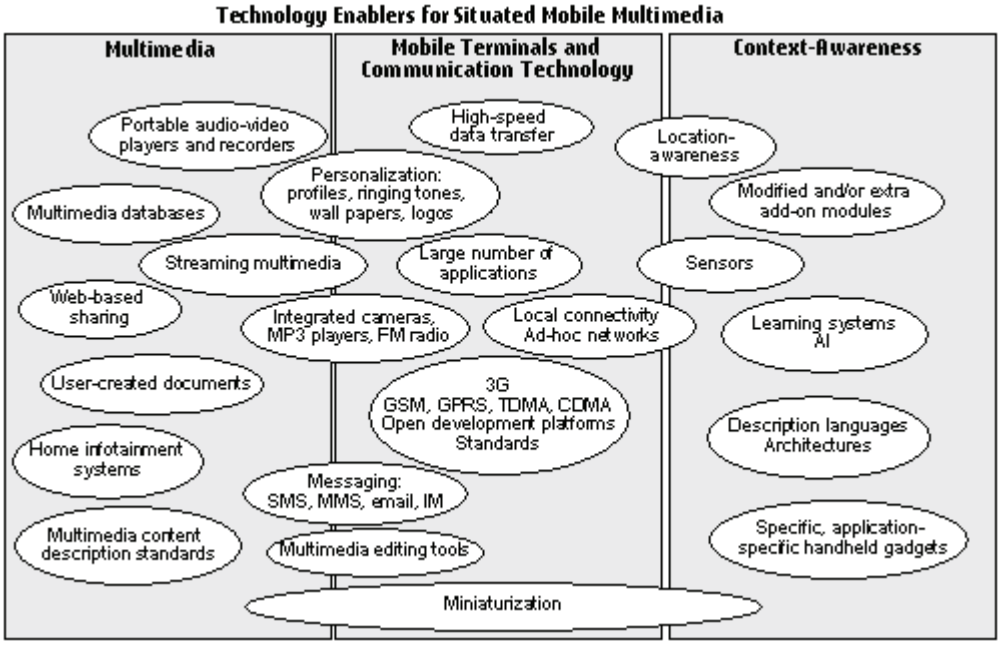
CURRENT STATUS

Technology Enablers

With the term "technology enabler," we mean the state-of-the-art technology that is mature enough and commonly available for building systems and applications based on it. When looking at the current status of the technology enablers for situated multimedia (Figure 3) it can be seen that there are several quite different factors related to the three domains of multimedia, mobile technology and context awareness that are still quite separated from each other. Recent developments have brought mobile technology and multimedia closer to each other, integrating them into mobile phone personalization, multimedia messaging, and imaging applications. Context awareness and mobile communication technology have moved closer to each other mainly on the hardware frontier, as GPS modules and integrated light-intensity sensors and accelerometers have been introduced for a number of mobile phones. Altogether, one can say that the areas presented in Figure 3 are under intensive development, and the features are overlapping more and more across the different domains. The development in hardware miniaturization, high-speed data transfer, data packaging, artificial intelligence, description languages, and standardization are all trends that lead toward seamless integration of the technology with versatile and robust applications, devices, and infrastructures.

A closer look at the current status of context awareness shows that the applications are typically built on specific gadgets, which include such device-specific features as modified hardware. However, the development is toward implementing the features on commonly used gadgets, such as off-the-self mobile phones or PDA's, so that platform-independent use of applications and services is possible. Also, applications utilizing location-awareness have typically concentrated on a defined, preset environment, where the beacons

Figure 3. Technology enablers and their current status for situated mobile multimedia



for location-detection have been placed across a limited, predefined environment, such as a university campus, distinct building, or certain area of the city. Thus, the infrastructure has not yet been generally utilized, and expanding it would require an extra effort. So far, the accuracy of GPS or cellular-ID-based recognition has often been too poor for experimenting with location-sensitive device features or applications requiring high location detection resolution. In mobile computing, it is challenging to capture a context, as a description of a current (e.g., physical situation) with a full confidence (==1). Various machine intelligence and data analysis-based methods such as self organizing neural networks (Flanagan, Mäntyjärvi, & Himberg, 2002), Bayesian approach (Korpipää, Koskinen, Peltola, Mäkelä, & Seppänen, 2003), fuzzy reasoning (Mäntyjärvi & Seppänen, 2003), and hidden Markov models (Brand, Oliver, & Pentland, 1997), to mention a few, have been studied to. In most approaches, quite good context recognition accuracies (~70-100%) are presented. However, it must be noted that all results are obtained with different and relatively limited data sets and results are quite subjective. The mobile context aware computing research, particularly context recognition, still lacks the systematic approach (e.g., benchmark datasets).

When examining the current technological status in mobile multimedia, the strongest trend during recent years has been the introduction of the multimedia messaging service (MMS), which has now established its status as an "everyday technology" with widespread use of so-called smart phones and camera phones. In Europe, it is estimated that people sent approximately 1.34 billion multimedia messages in 2005. This shows that MMS is a considerable potential technology that end users have become to adopt, although this is only a small fraction of the number of SMSs sent, which has been estimated to be 134,39 billion in 2005 (Cremers & de Lussanet, 2005).

Personalization of mobile phones, which so far has been executed manually by the user, has taken the form of changing ringing tones, operator

logos, and wallpapers. Multimedia offers further possibilities for enhancing the personalization of the mobile device — both from the user's self-expression point of view when creating his or her own items and as the receiving party when the user may access multimedia content via peer-to-peer sharing, information delivery, or mobile services.

Toward Situated Mobile Multimedia

The research in the area of situated mobile multimedia is still in its early development stage and many of the current projects are very limited, still concentrating mainly on textual information exchange. The most common contextual information source used in mobile communication is the location. In addition to the information bound to the physical location, information on the current physical location and distance may provide useful data for (e.g., time management and social navigation). E-graffiti introduces an on-campus location-aware messaging application where users can create and access location-associated notes, and where the system employs laptop computers and wireless network-based location detection (Burrell & Gay, 2002). InfoRadar supports public and group messaging as a PDA application, where the user interface displays location-based messages in a radar-type view showing their orientation and distance from the user (Rantanen, Rantanen, Oulasvirta, Blom, Tiitta, & Mäntylä, 2004). The applications exploiting multimedia elements are typically location-based museum or city tour guides for tourists (see e.g., Davies & al., 2001).

Multimedia messaging has become a popular technique for experimentation within the field since there is no need to set up any specific infrastructure and standard mobile phones can be used as the platform. The widespread use of mobile phones also enables extending the experiments to large audiences, as no specific gadgets need to be distributed. These aspects are used in the work of Koch and Sonenberg (2004) for developing an MMS-based location-sensitive museum information application utilizing Bluetooth as the sensing technology. In the Rotuaari project carried out in Oulu, Finland, location-aware information and advertisements were delivered to mobile phones in the city center area by using different messaging applications, including MMS ("Rotuaari," n.d.).

Use of context can be divided into two main categories: push and pull. In the push type of use, the context information is used for automatically sending a message to the user when he or she arrives at a certain area, whereas with the pull type, the user takes the initiative by requesting context-based information, such as recommended restaurants in the prevailing area.

Currently, most of the experiments concentrate on the push type of service behavior. This is partially due to the lack of general services and databases, which, in practice, prevents the use of the request-based approach. A general problem is the shortage of infrastructure supporting sensing, and the absence of commonly agreed principles for service development. Attempts to develop a common framework to enable cross-platform application development and seamless interoperability exist, but so far there is no commonly agreed ontology or standard.

In Häkkilä and Mäntyjärvi (2005) we have presented a model for situated multimedia and how context-sensitive mobile multimedia services could be set, stored, and received, and examined users' experiences on situated multimedia messaging. The model combines multimedia messaging with the context awareness of a mobile phone and phone applications categorized into three main groups, notification, reminder, and presence, which were seen to form a representative group of applications relevant to a mobile handset user. The composing entity (i.e., a person or a service) combines the device application information, the multimedia document, and the context information used for determining the message delivery

conditions to a situated multimedia message. After sending, the message goes through a server containing the storage and context inferring logic. The server delivers the message to the receiving device, where it has to pass a filter before the user is notified of the received message. The filter component prevents the user from so-called spam messages and enables personalized interest profiles.

FUTURE TRENDS

In order to successfully bring new technology to the use of a wide audience, several conditions must be fulfilled. As discussed before, the technology must be mature enough so that durable and robust solutions can be provided at a reasonable price, and an infrastructure must be ready to support the features introduced. The proposed technological solutions must meet the end users' needs, and the application design has to come up with usable and intuitive user interfaces in order to deliver the benefits to the users. Importantly, usable development environments for developers must exist. Usability is a key element for positive user experience. In the ISO 13407 (3.3), standard on human-centred design processes for interactive systems, usability has been defined to be the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (Standard ISO 13407, 1999).

In this section, we discuss the near and medium-term future concepts focusing on situated mobile multimedia enabled by the development trends in context awareness, multimedia technologies, and mobile terminals.

Context Awareness

Context awareness has been recognized as one of the important technology strategies on the EC level (ITEA, 2004). The key factors for the human-system interaction of mobile devices recognized by ITEA are: simple, self-explanatory, easy to use, intelligent, context-aware, adaptive, seamless, and interoperable behaviour of user interfaces. The main driving forces for the development of context-awareness for mobile devices are that the user interface of mobile terminals is limited (by the small physical size), and there is an enormous expected growth in mobile applications and services to be accessed by terminals in near future. So, it can be said that the need for context-awareness is evident and the need is strongly market and industry-driven. The main enablers for the context-awareness of mobile devices are: ease of use and an open development environment for smart phones, architectures for mobile context-awareness enabling advanced context reasoning, miniaturized and low-power sensing technologies and data processing, and suitable languages for flexibly describing the context information. These main enablers are recognized in the research field and the research is rapidly advancing globally toward true context-awareness.

However, promising technological solutions often have some drawbacks. In context awareness, they are related to user experience. While the nice scenarios of context awareness provide a better tomorrow via intelligently behaving user interfaces—the right information in the right situation—the reality might be a bit different. For example, entering a new context may cause the adapted UI to differ radically from what it was a moment ago, and a user may find him or herself lost in the UI. On the other hand, the device may incorrectly recognize the situation and behave in an unsuitable manner. These are just examples of horror usability scenarios, but the idea is that the responsibility for the functionality of the device is taken away from the user, and when a user's experience of a product is negative, the consequences for the terminal business may be fatal.

There are ways to overcome this problem. One—and a careful—step toward successful context-awareness for mobile terminals is to

equip the devices with all the capabilities for full context-awareness and provide an application, a tool by which a user him or herself may configure a device to operate in a context-aware manner (Mäntyjärvi et al., 2003). Obviously, this is an opportunity to accomplish context-aware behavior, but on the other hand, the approach sets more stress on the user when he or she must act as an engineer (so-called end user programming), and continuous configuration may become a nuisance. However, this approach is more attractive for the terminal business when the user him or herself is responsible for any possible unwanted behavior of a device instead of the device manufacturers.

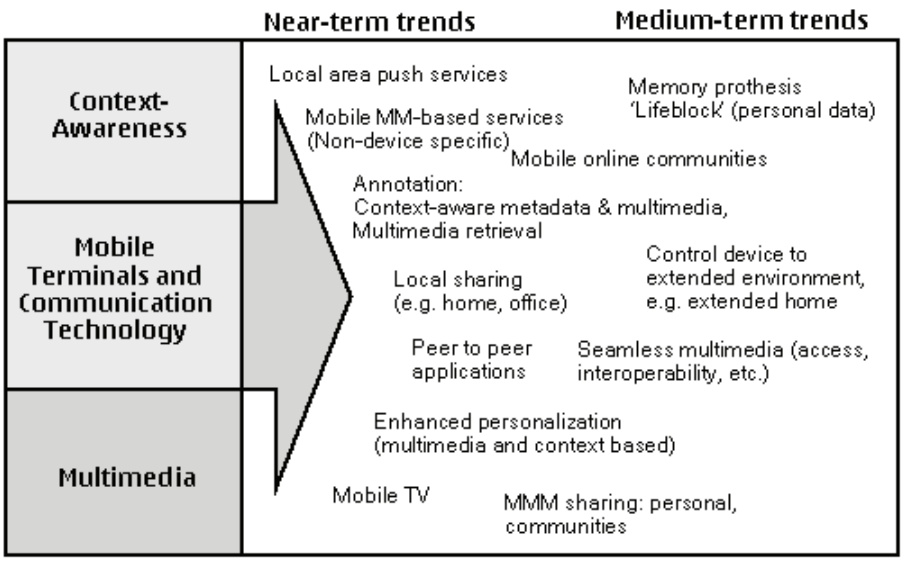
Future Development Trends

In the very near future, we are to witness a growth in basic wireless multimedia applications and services (Figure 4). By basic services and applications we refer to local services (push multimedia kiosks, mobile mm content, etc) and mobile multimedia services, including downloadable content: ringtones, videos, skins, games, and mobile terminal TV. Streaming multimedia is strongly emerging, and mobile TV in particular has raised expectations as the next expanding application area. At the moment, several broadcasting standards are dominating the development in different parts of the world: ISDB-T in Japan; DVB-H, particularly in Europe and the US; and DBM, especially in Korea, where several mobile terminals on the market already support it. Although the status of mobile TV is not reaching a wide audience of end users yet, a number of trials are in existence, such as a half-year trial in Oxford (UK) with NTL Broadcast and O2 starting in the summer of 2005, when 350 users will access several TV channels with the Nokia 7710 handset (BBC News, 2005).

The increased growth in MM services is supported by the maturation of various technologies, including 3G-networks, extensive content communication standards such as MPEG 7, multimedia players and multimedia editing tools for mobile terminals. In addition, an increase in the amount of mobile multimedia is expected to support the stimulation of the mobile service business.

The increase in the amount of personal multimedia is expected to be considerable, mainly

Figure 4. Future development trends enabled by context-awareness, mobile terminals, and communications technology and multimedia



because of digital cameras and camera phones. The explosion of mobile personal multimedia has already created an evident need for physically storing data—by which we mean forms of memory: memory sticks and cards, hard-drive discs, etc.

However, another evident need is end user tools and software for managing multimedia content—digital multimedia albums, which are already appearing in the market in the form of home multimedia servers enabling communication of multimedia wirelessly locally at home, and personal mobile terminal MM album applications.

The next overlapping step in the chain of development is the mobile online sharing of the multimedia content of personal albums. People have always had a need to get together (i.e., the need to form communities) (e.g., around hobbies, families, work, etc). Today, these communities are in the Web and the interaction is globally online. Tomorrow, the communities will be mobile online communities that collaborate and share multimedia content online.

In the near future, as the personal multimedia content is generated with context-aware mobile phones, the content will be enhanced semantic context-based annotations, with time, place, and social and physical situation. The created metadata will be sketched with effective description languages enabling more effective information retrieval in the mobile semantic web and more easily accessible and easy to use multimedia databases.

Even though we have only identified a few near-term concepts enabled by the combination of context awareness and mobile multimedia technologies, we can also see the effects in the long term. The role of the mobile terminal in creating, editing, controlling, and accessing personal and shared multimedia will emphasize. The emerging standards in communication and in describing content and metadata enable seamless access and interoperability between multimedia albums in various types of systems. The personal "My Life" albums describing the entire life of a person in a semantically annotated form are become commonplace.

CONCLUSION

In this chapter we have examined the concept of situated mobile multimedia for mobile communications. We have introduced the characteristics of the key elements — multimedia, context awareness, and mobile communications — and discussed their current status and future trends in relation to the topic. Linked to this, we have presented a new categorization for contextual information sources, taking account of the special characteristics of mobile communication devices and their usage (Figure 2).

Integrating context awareness into mobile terminals has been introduced as a potential future technology in several forums. The motivation for this arises from the mobility characteristics of the device, its limited input and output functionalities, and the fact that the complexity of the device and its user interface is constantly growing. Context awareness as such has potential to functions such as automated or semi-automated action executions, creating shortcuts to applications and device functions, and situation-dependent information and service delivery. In this chapter we have limited our interest to examining the possibilities of combining the multimedia content to the phenomenon.

Currently, most of the mobile devices employing context awareness are specifically designed for the purpose, being somehow modified from the standard products by adding sensor modules. The lack of commonly agreed ontologies, standards, and description languages, as well as the shortage of suitable, commonly used gadgets as application platforms, has hindered the development of generally available, wide-audience services, and applications.

Multimedia, especially in the form of camera phones and Multimedia Messaging Service, has become a solid part of mobile communication technology during recent years. MMS enables easy delivery of multimedia content to a broad audience in a personalized manner. With situated multimedia, this means information delivery with multimedia content, such as information on local services or current events. Context awareness can also be exploited for executing underlying actions hidden from the user, such as selecting the data transfer medium for lower-cost streaming or better connection coverage.

REFERENCES

BBC News. (2005, May 11*). Mobile TV tests cartoons and news.* Retrieved June 15, 2005, from http://news.bbc.co.uk/1/hi/technology/4533205.stm

Brand, M., Oliver, N., & Pentland, A. (1997). Coupled hidden Markov models for complex action recognition. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition.*

Burrell, J., & Gay, G. K. (2002). E-graffiti: Evaluating real-world use of a context-aware system. *Interacting with Computers, 14*(4), 301-312.

Cole, H., & Stanton, D. (2003). Designing mobile technologies to support co-present collaboration. *Personal and Ubiquitous Computing, 7*(6), 365-371.

Cremers, I., & de Lussanet, M. (2005, March). *Mobile messaging forecast Europe: 2005 to 2010.* Cambridge, MA: Forrester.

Davies, N., Cheverst, K., Mitchell, K., & Efrat, A. (2001). Using and determining location in a context-sensitive tour guide. *IEEE Computer 34*(8), 35-41.

Dey, A. K., & Abowd, G. D. (2000). Toward a better understanding of context and context-awareness. In the *CHI 2000 Workshop on the What, Who, Where, When, Why, and How of Context-Awareness.*

Flanagan, J., Mäntyjärvi, J., & Himberg, J. (2002). Unsupervised clustering of symbol strings and context recognition. In *Proceedings of the IEEE International Conference of Data Mining 2002* (pp. 171-178).

Gellersen, H.W., Schmidt, A., & Beigl, M. (2002). Multi-sensor context-awareness in mobile devices and smart artefacts. *Mobile Networks and Applications, 7*(5), 341-351.

Grinter, R.E., & Eldridge, M. (2003). Wan2tlk?: Everyday text messaging. *CHI Letters, 5*(1), 441-448.

Häkkilä, J., Beekhuyzen, J., & von Hellens, L. (2004). Integrating mobile communication technologies in a student mentoring program. In *Proceedings of the IADIS International Conference of Applied Computing 2004* (pp. 229-233).

Häkkilä, J., & Mäntyjärvi, J. (2005). Combining location-aware mobile phone applications and multimedia messaging. *Journal of Mobile Multimedia, 1*(1), 18-32.

ITEA Technology Roadmap for Software-Intensive Systems (2nd ed.). (2004). Retrieved June 15, 2005, from http://www.itea-office.org/newsroom/publications/rm2_download1.htm

Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005, April-June). The ubiquitous camera: An in-depth study on camera phone use. *Pervasive Computing, 4*(2), 42-50.

Koch, F., & Sonenberg, L. (2004). Using multimedia content in intelligent mobile services. In *Proceedings of the WebMedia & LA-Web 2004* (pp. 41-43).

Korpipää, P., Koskinen, M., Peltola, J., Mäkelä, S. M., & Seppänen, T. (2003). Bayesian approach to sensor-based context-awareness. *Personal and Ubiquitous Computing, 7*(2), 113-124.

Kurvinen, E. (2003). Only when Miss Universe snatches me: Teasing in MMS messaging. In *Proceedings of DPPI'03* (pp. 98-102).

Mäntyjärvi, J., & Seppänen, T. (2003). Adapting applications in mobile terminals using fuzzy context information. *Interacting with Computers, 15*(4), 521-538.

Mäntyjärvi, J., Tuomela U., Känsälä, I., & Häkkilä, J. (2003). Context studio–tool for personalizing context-aware applications in mobile terminals. In *Proceedings of OZCHI 2003* (pp. 64-73).

Rantanen, M., Oulasvirta, A., Blom, J., Tiitta, S., & Mäntylä, M. (2004). InfoRadar: Group and public messaging in the mobile context. In *Proceedings of NordiCHI 2004* (pp. 131-140).

Rotuaari. (n.d.) *Rotuaari*. Retrieved June 15, 2005, from http://www.rotuaari.net/?lang=en

Schilit, B., Adams, N., & Want, R. (1994) Context-aware computing applications. In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications* (pp. 85-90).

Schmidt, A., Beigl, M., & Gellersen, H. (1999). There is more context than location. *Computers and Graphics Journal, 23*(6), 893-902.

Standard ISO 13407. (1999). *Human-centred design processes for interactive systems.*

Tachikawa, K. (2003, October). A perspective on the evolution of mobile communications. *IEEE Communication Magazine, 41*(10), 66-73.

KEY TERMS

Camera Phone: Mobile phone employing an integrated digital camera.

Context Awareness: Characteristic of a device that is, to some extend, aware of its surroundings and the usage situations

Location Awareness: Characteristic of a device that is aware of its current location.

Multimedia Messaging Service (MMS): Mobile communication standard for exchanging text, graphical, and audio-video material. The feature is commonly included in so-called camera phones.

Situated Mobile Multimedia: Technology feature integrating mobile technologies, multimedia, and context awareness.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 326-339, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.14
Current Status of Mobile Wireless Technology and Digital Multimedia Broadcasting*

J. P. Shim
Mississippi State University, USA

Kyungmo Ahn
Kyunghee University, Korea

Julie M. Shim
Soldier Design LLC, USA

ABSTRACT

The purpose of this chapter is to present an overview of wireless mobile technology, its applications, with a focus on digital multimedia broadcasting (DMB) technology. The chapter also explores the research methodology regarding users' perception on DMB cellular phones and presents empirical findings. Implications for future research are presented. The report attempts to provide stimulating answers by investigating the following questions: (1) Do users perceive easy access to DMB applications as a satisfactory service offered by DMB service providers? (2) Do users perceive high-quality DMB program content as a satisfactory service offered by the DMB service providers? (3) Are there differences between different age groups in terms of their perception of DMB phone prices, phone usage time, program content, and services?

INTRODUCTION

Wireless mobile technology and handheld devices are dramatically changing the degrees of interaction throughout the world, further creating a ubiquitous network society. The emergence of these wireless devices has increased accuracy, ease-of-use, and access rate, all of which is increasingly essential as the volume of information handled by users expands at an accelerated pace.

Mobile TV broadcasting technology, as a nascent industry, has been paving a new way to create an intersection of telecommunication and media industries, all of which offers new opportunities to device makers, content producers, and mobile network operators.

There are currently various wireless connectivity standards (e.g., Wi-Fi, Bluetooth, Radio Frequency Identification [RFID], etc.), which have been expanding across all vertical industries, in an era of mobile and ubiquitous computing, which provides access to anything, anytime, and anywhere. Mobile TV technologies have been creating a buzz, as it adds a new dimension to the "on the go" mobility factor—simultaneous audio and video services are broadcasted in real-time to mobile devices in motion, such as mobile TV-enabled phones, PDAs, and car receivers.

There are currently three major competing standards: digital video broadcasting for handhelds (DVB-H), which is going through trial phases in Europe; digital multimedia broadcasting (DMB), which has been adopted in South Korea and Japan; and MediaFLO (QUALCOMM Inc., 2005), which is currently in trial phase in the United States with plans to launch by late 2007. The competition scheme is further intensified given the challenge of how quickly terrestrial and satellite DMB can be deployed and commercialized throughout countries such as Korea, Japan, and Europe. Additionally, there is pressure to recoup the costs with creating the network and catapult the technology to the ranks of industry standard.

The purpose of this chapter is to present an overview of wireless mobile technology, its applications, with a focus on DMB technology. The chapter also explores the research methodology regarding users' perception on DMB cellular phones and presents empirical findings from Study Phases I and II, along with actual DMB subscriber usage results. Implications for future research are presented.

Given that the research topic of DMB has not yet been covered extensively, the use of qualitative methods is considered advantageous when exploring the topic to develop theoretical variables, which may then be employed in quantitative research. Thus, with the difference found between the DMB cellular phone usage experience and traditional cellular phone usage, qualitative methodology was applied to Study Phase I. The project was then triangulated by the use of quantitative methodology in Study Phase II to develop an additional understanding of the DMB cellular phone users' experiences as identified in Study Phase I.

The report attempts to provide stimulating answers by investigating the following questions: (1) Do users perceive easy access to DMB applications as a satisfactory service offered by DMB service providers? (2) Do users perceive high-quality DMB program contents as a satisfactory service offered by the DMB service providers? (3) Are there differences between different age groups in terms of their perception of DMB phone prices, phone usage time, program contents, and services?

WIRELESS MOBILE TECHNOLOGIES: CURRENT STATUS AND CONCEPTS

Over the last decade, wireless technologies have attracted unprecedented attention from wireless service providers, developers, vendors, and users. These wireless technologies provide many connection points to the Internet between mobile phones and other portable handheld devices to earpieces and handsets. These technologies include Wi-Fi hotspots, Bluetooth, WiMAX, wireless broadband Internet (WiBRO), RFID, and others. Wi-Fi hotspots, with a distance and penetration of approximately 50 feet, are physical addresses where people can connect to a public wireless network, such as a cafe, hotel, or airport. WiMAX

is a metropolitan-scale wireless technology with speeds over 1Mbps and a longer range than Wi-Fi. WiBRO, the Korean version of WiMAX, allows users to be connected to the Internet while in motion, even in cars traveling up to 100 kilometers per hour. It is anticipated that users may one day seamlessly switch between networks multiple times per day, depending on the service offered by a specific network service provider.

Many industries have seen the benefits of these wireless technology applications, of which some will be described here. For local, federal, and state agencies, wireless connections provide for GPS functionality, along with real-time vehicle tracking, navigation, and fleet management. For automated logistics and retail industries, RFID tags will give information on just-in-time inventory or shipment location, security status, and even environmental conditions inside the freight. In the health care industry, the wireless applications include patient and equipment monitoring, and telemedicine through the monitoring of an outpatient's heart via continuous electrocardiograms (ECG). Other applications already on the radar: handsets that function as a blood pressure monitor, a blood glucose meter, and wireless pacemaker. One of the hurdles that wireless solution carriers have to overcome is the cost of the devices, and whether insurance companies are willing to cover or share the costs. The wireless technology allows government officials and emergency response teams to stay informed of critical information in the event of an emergency or a disaster that affects wire line services, much like Katrina; these include advanced warnings and public alerts, emergency telecommunications services, global monitoring for environment, and assistance with search and rescue (SAR).

PC World, an online technology magazine, recently reported that the number of Wi-Fi hotspots reached the 100,000 mark globally.[1] Businesses are realizing the value-added service by offering free or paid wireless services to attract customers. Analysts believe that locations such as school campuses and citywide deployment of WiMax technology will benefit users.

a. **United States:** Wi-Fi integration into retail, hospitality, restaurant, and tourism industries has been instrumental for marketing plans, particularly for franchise venues, including Starbucks and McDonald's.
b. **Asia/Pacific:** An article in *The Australian* (2003, March 4) described that 200 restaurants in Australia have migrated away from taking orders via pen and paper to using wireless handhelds to relay orders to the kitchen/bar staff. In addition to offering this type of service, Japan's NTT DoCoMo introduced its iMode Felica handset, enabling users to scan their handsets as their mobile wallets (m-wallets), eliminating the need to carry a credit card, identification, and keys. The feature allows for conducting financial transactions, purchasing services/products, or opening electronic locks.[2] The issue at hand is the different business models of the wireless carrier and that of the credit card companies.

DIGITAL MULTIMEDIA BROADCASTING: CURRENT STATUS AND CONCEPTS

Digital multimedia broadcasting (DMB) is a process of broadcasting multimedia over the Internet or satellite that can be tuned in by multimedia players, capable of playing back the multimedia program.[3] DMB is an extension of digital audio broadcasting (DAB), which is based on the European Eureka 147 DAB Radio standard. DMB technology has two sub-standards: satellite-DMB [S-DMB] and terrestrial-DMB [T-DMB]. While both S-DMB and T-DMB broadcasts television to handheld devices in motion, the difference lies in the transmission method: via satellite versus land-based towers. These real-time transmissions

allow users to view live TV programs, including news, reality shows, or sports games on their DMB cellular phones in the subway.

With mobile growth two or three times that of Europe and North America (Budde, 2002), Japan and Korea have been known for their cutting edge technological innovations and tech-savvy consumers. Korea is one of the world's most broadband-connected countries, with a high penetration rate (Lee, 2003; Shim et al., 2006a; Shim et al., 2006b). The government initiatives have been instrumental in this arena, as the government's hands-on style has created the IT infrastructure necessary to power the latest technological tools. The mobile markets in Japan and Korea have become optimal testing grounds for mobile operators and manufacturers before rolling out products in the rest of the world, given the consumers' insatiable appetite of acquiring the latest technologies, early acceptance behavior, and education fever.

In Asia-Pacific and Europe, considered to be the power houses of the mobile gaming industry, wireless gaming and instant messages have exceeded expectations. In North America, music downloads and e-mails have become essential. As the market for mobile applications, (including short message service [SMS], ring-tones, games, music, videos) is becoming more saturated, more wireless applications have become integrated into most consumer electronics devices, from digital cameras to video game consoles. With over 85% cellular phone penetration rate, Korea introduced the world's first DMB mobile-enabled phone, or "TV-on-the-go" in 2005.[4] While Japan currently provides S-DMB services designed for car receivers, Korea has been the only country to provide full-blown S-DMB and T-DMB services on cellular phones while in motion (including car receivers) by late 2006. With T-DMB and S-DMB services already launched in Korea, several countries in Europe, and the U.S. are planning to launch DVB-H services by the end of 2007. Informa, a consultancy, says there will be 125 million mobile TV users by 2010.[5]

The history of DMB began with the development of DAB services during the mid-1990s in the U.S. and Europe (Korean Society for Journalism and Communication Studies, 2003; Nyberg, 2004). The current status of Mobile TV services in the U.S., Europe, Japan, and Korea is shown in Table 1 (Shim, 2005b).

Table 1. Current status of mobile TV services in various countries

Country	USA		Europe		Japan	Korea	
Mobile TV technology	MediaFlo	DVB-H	DVB-H	T-DMB	S-DMB	S-DMB	T-DMB
Receiving device	Car receiver	Car receiver	Mobile TV- phone, Car receiver	Car receiver	Car receiver	Mobile TV- phone, Car receiver	Car receivers
Service launch date	2006	2006	2006	2006	2004	May 2005	Dec 2005

Sources: The Korea Times, (2005, January 18) "Korea's Free Mobile Broadcasting Faces Snag".
KORA Research 2003-10,. (2004, May). "A Market Policy Study on DMB".
M. H. Eom, "T-DMB Overview in Korea," (2006, April). Proceedings of 2006 Wireless Telecommunications Symposium, Pomona, CA,

As shown in Figure 1, DMB program producers provide a variety of programs and content to the DMB center, which broadcasts through either satellites or towers. Thus, the DMB cellular phone users receive content and programs through satellites, towers, or "gap-fillers" (small base stations) to ensure there are no reception problems, even in underground subways (Shim, 2005a).

Consumers are increasingly gravitating towards customized devices and features, as a miniaturized interactive entertainment center is packaged into the cellular phone, complete with an MP3 player, multi-megapixel camera, digital video recorder, CD-quality audio, and a selection of satellite broadcast television and audio channels (Olla & Atkinson, 2004) as they can choose from television and audio on-demand and simultaneously make phone calls. The mobile TV-enabled phone, equipped with these features, has become more than integrated into one's lifestyle, as it becomes an extension of the consumer's identity. The handset carriers are in the process of yet again trying to capitalize on producing fashion-forward phones and portable gaming consoles.

DMB data service is a framework of the following groups: data provider, audio/video content producer, DMB producer, advertiser, and customer. A schematic view of DMB data service and the components, shown in Figure 2, provides a basic understanding of the general structure of the DMB business model. The figure also shows interaction of the DMB producer with other groups of DMB data services.

For example, the DMB producer provides various content and programs to customers for a service fee. The DMB producer charges an advertising fee to the advertiser, from whom customers can purchase directly for advertised services via the DMB device. The audio/video content producer and data provider each provide various contents to the DMB producer for a fee. The perceived richness of the medium should have an impact on the use of the communication medium (Daft & Lengel, 1986; Smagt, 2000). The rich media is more appropriate in ambiguous communications situations, which emphasizes Daft and Lengel's valuable contribution of placing equivocality high in the business and information systems field.

Figure 1. An overview of the mobile wireless framework

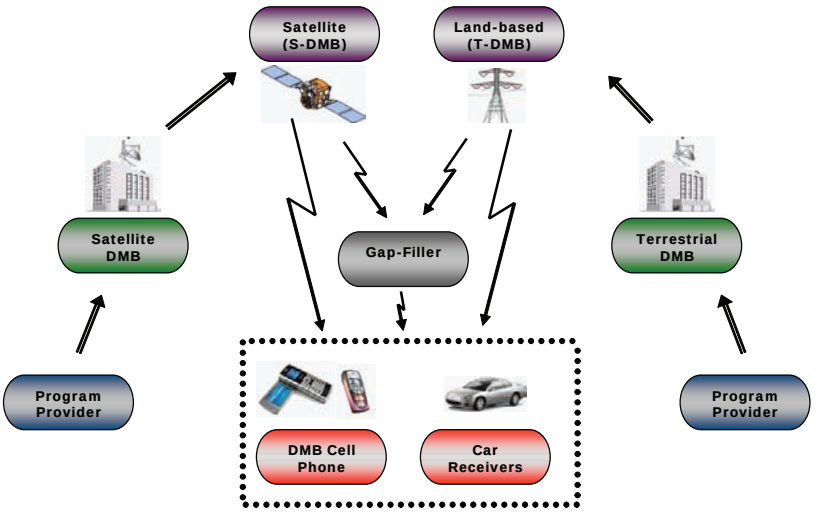
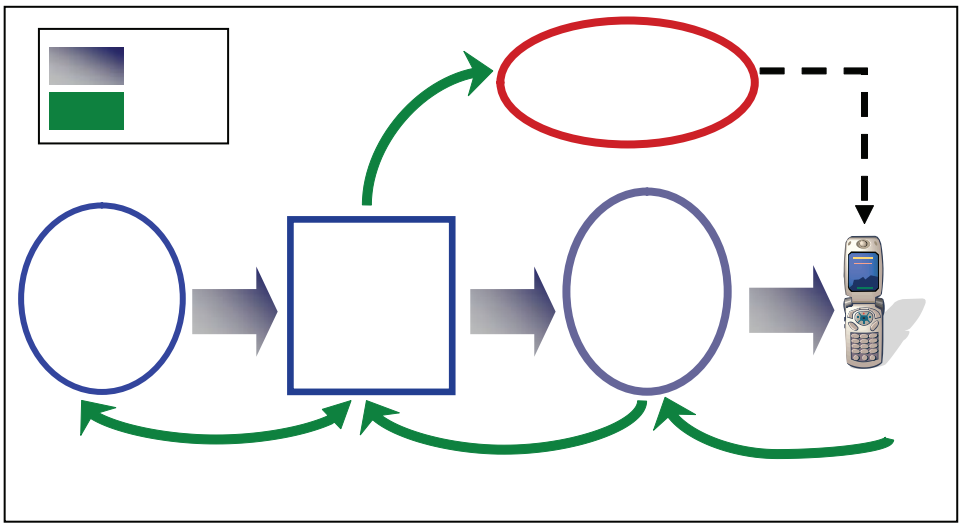


Figure 2. A schematic view of DMB data service business model



Source: Modified from KORA Research 2003-10. (2004, May). "A Market Policy Study on DMB," Research Report of Korea Radio Station Management Agency.

There exists a rich body of knowledge of technology adoption and diffusion, including the digital multimedia broadcasting technology. For example, several theoretical backgrounds, such as institutional theory, technology acceptance model (TAM) (Venkatesh & Davis, 2000), and diffusion of innovation theory (Gharavi, Love, & Cheng, 2004; Rogers, 1983) explain the DMB technology adoption at an individual, organizational, and industry level (Lee, 2003; Shim, 2005a, 2005b). Among the theories, Lee and Shim both describe the major factors behind Korea's information and communication technology diffusion such as: external factors (global economy, government policies), innovation factors (usefulness, ease of use, self-efficacy), and imitation factors (subjective norm of belongingness, word of mouth). The authors believe that either the diffusion theory (such as external, internal, and mixed influence models), or TAM (such as perceived usefulness and perceived ease of use), or the combination of both can be applied behind DMB cellular phone adoption and diffusion.

RESEARCH METHODOLOGY

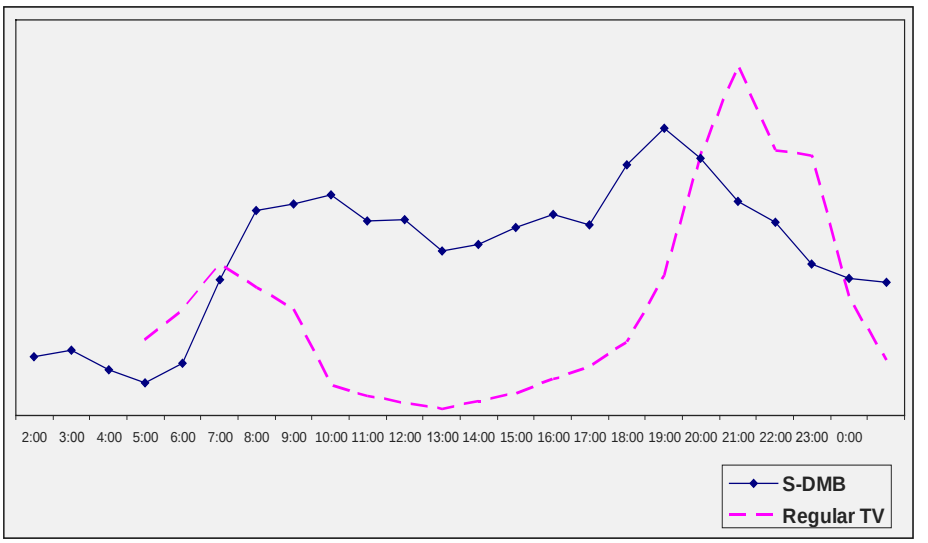
A recent study demonstrates a higher number of DMB viewers than regular TV viewers during the daytime (Figure 3). Since the DMB cellular phone captures the content-on-demand aspect, the DMB phone service (S-DMB) are optimal for the on-the-go daytime enthusiasts.

To determine how integral DMB phones have been and will be in consumers' daily lives, the authors conducted qualitative and quantitative analyses. Study Phase I describes the use of the qualitative research method, specifically the existential phenomenological method. Study Phase II describes the quantitative research methods including the survey questionnaire (Shim, Shin, & Nottingham, 2002).

STUDY PHASE I: QUALITATIVE ANALYSIS

Although quantitative instruments serve as valid methods to study the perceived use of DMB

Figure 3. The percentage of viewing on S-DMB vs. regular TV



Source: Suh, Y. (2005, November). "Current Overview of S-DMB," TU Media.

phones, qualitative research methods, such as interviewing, can reveal the function of variables perhaps overlooked by survey designers. The current project was designed to employ the qualitative technique of existential phenomenology. Thus, it develops an in-depth understanding of the new concept of DMB usage by investigating respondents' reports of their DMB phone usage experiences. With this data collection technique, the respondent is encouraged to describe in-depth the personally experienced phenomenon (Thompson, Locander, & Pollio, 1989).

Existential phenomenology was selected among various qualitative methods, such as case studies and ethnography, because of its attention to a respondent's individualistic, subjective expression of an actual live experience of the situation of interest. Such reflection on a single experience encourages the perceiver to focus on nuances that would likely escape the broader brush of a researcher's selection of choices among a pre-set list of quantitative dimensions or escape even the surface comparison of reports of respondents' experiences. Existential phenomenology encourages the respondent to consider specific and live events. The goal is to discover patterns of experiences (Thompson et al., 1989).

Since the purpose of existential phenomenology is to describe the experience as it is lived, the interview has been found to be a powerful tool for attaining in-depth understanding of another person's experience (Kvale, 1983). Research analysis of interview-derived information is considered valid because the respondents' own words are used to understand their experiences (Feagin, Orum, & Sjoberg, 1991). Accordingly, respondents in this research were presented with a set of open-ended questions designed to encourage them to discuss and describe their experiences with DMB phone usage. To determine a specific set of key factors that would be of critical concern to DMB users, 19 respondents in Korea were enlisted in Study Phase I.

A purposive sample is deemed appropriate for exploratory research designed to query respondents who have experienced a phenomenon of interest. Thus, the networking technique was utilized to obtain a purposive sample of individuals who had interacted with the DMB cellular phone services. These respondents were then

asked to name additional individuals who had experienced DMB services. Thus, aside from the requirement of the respondents' familiarity with the DMB services, demographic characteristics of the sample resulted by random chance.

The majority of respondents were well-educated young professionals with a zealous tech-gadget nature, affluent, computer proficient, and somewhat knowledgeable about DMB services. Although this sample clearly is not representative of the population at large, the sample profile corresponds with what the authors presumed to be identified as a typical DMB service user. Thus, the experiences relayed by these respondents are considered to be a reasonable representation of a random sampling of regular DMB cellular phone users. After respondents were assured of confidentiality and protection of their privacy, each tape-recorded interview lasted 20-30 minutes. Each interview began with open-ended questions posed in a conversational format to encourage the respondent to develop a dialogue with the interviewer, providing the context from which the respondent's descriptions of his or her own DMB service experience could flow freely and in detail. Participants were encouraged to discuss not only their DMB services experiences, but also their attitudes and perceptions regarding negative and positive aspects of DMB services.

Such in-depth descriptions have been found to be beneficial in revealing emotional and behavioral underpinnings of overt user behavior. In reality, the act of a respondent's description of a specific experience in-depth, frequently results in further personal insights that arise through the revival of the experience. The respondents were asked to describe their main reasons for purchasing DMB cellular phones, which varied: "to gain information access," "to spot the latest trends," "for education or entertainment," "to watch TV while commuting," and "for movies, dramas, and shopping." Their personal positive experiences were: "mobility—a deviation from a fixed location point," "high quality reception," "convenience," "accessible anytime/anywhere," "lifestyle change," "great for commuting," and "good for managing time." On the other hand, the negative aspects they experienced included: "expensive device," "reception problem," "low battery hours with limited usage time (e.g., 2-3 hours)." Most respondents reported that the following areas would have great potential for future DMB applications and content: information access, education/learning, e-trading, retail, tourism, and entertainment. In Study Phase II, these themes were reconstructed to set up independent variables for the quantitative analysis.

STUDY PHASE II: QUANTITATIVE ANALYSIS

To determine the extent to which DMB phones are being used as the latest multimedia product, the authors developed a questionnaire. DMB has a wide array of advantages: personalized, live media (television, radio, or data broadcasts) that can be viewed on-demand anytime; the mobility of the phone which receives satellite and terrestrial television broadcast signals even at high speeds or underground; and an interactive handset into which one can speak via the handset while watching TV programs. The research instrument underwent two pretests. The first pretest involved administering the questionnaire to 25 graduate and undergraduate students at a large university in Seoul, Korea. The questions, which concerned price, usage time, program content, and services were modified to reduce the effects of proximity bias on the responses, with several questions reworded for clarity. The second pretest was conducted at a DMB phone service provider company to ensure the content validity. A five-point Likert scale was used for recording the responses.

DMB will not be successful if content and service providers fail to provide high quality service, a variety of content, and reasonable prices for services and handsets (Teng, 2005).

Several research studies demonstrated that there are differences among age groups on factors such as technology adoption and usage (Larsen & Sorebo, 2005; Ventatesh, 2000). It is believed that older generations are more anxious about the use of technologies than the younger generations. A number of research studies have supported this belief (Gilbert, Lee-Kelley, & Barton, 2003). Based on the theories and research questions along with Study Phases I and II, the authors developed the following six hypotheses:

H_1: The user's easy access to DMB service is perceived as a satisfactory service offered by the DMB service provider.

H_2: Premium (excellent) content of DMB programs corresponds with a good quality DMB service provider.

H_3: There is a difference between different age groups and their perceived value of DMB handset price.

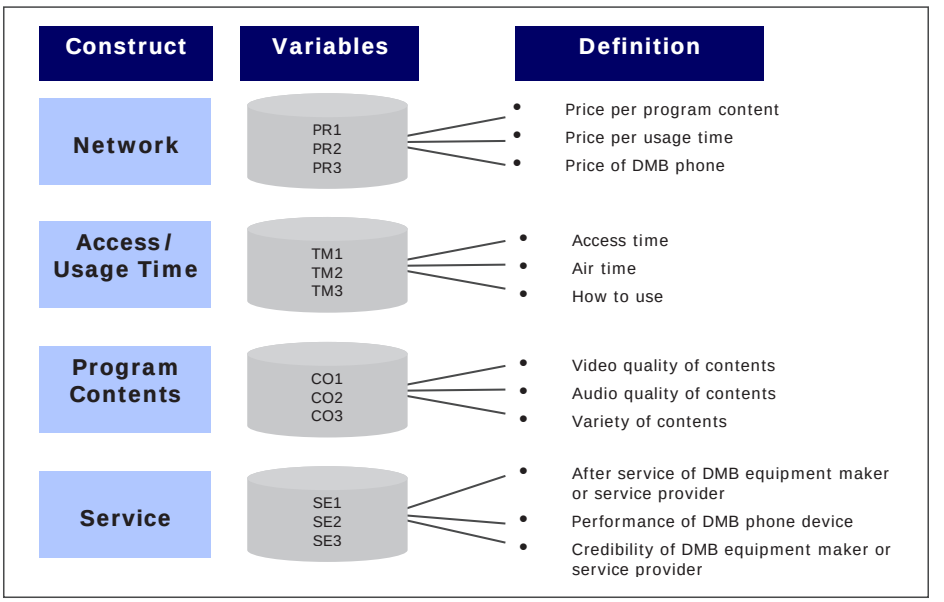
H_4: There is a difference between different age groups and their perceived value of DMB phone usage time.

H_5: There is a difference between different age groups and their perceived value of DMB program content.

H_6: There is a difference between different age groups and their perceived value of DMB services.

The authors and their research assistants distributed the questionnaire to 300 randomly selected individuals inside the Korea Convention Exhibition Center (COEX) and Korea World Trade Center during January and February 2005. Of the 300 randomly selected individuals' responses, 264 were valid. The two-page questionnaire was divided into three sections with a total of 32 questions. In Section 1, the authors asked the randomly selected participants about DMB services, such as information sources about DMB services, user satisfaction ratings, influential factors when choosing DMB services, DMB applications, and others. The questions in Section 2 covered the participants' perceived values of DMB application services. Section 3 inquired of participants' demographics.

Table 2. Model construct



DATA ANALYSIS AND FINDINGS

The 264 usable research instruments collected from the respondents were well represented in terms of gender, age, and occupation. Statistical Package for the Social Sciences (SPSS) was used to calculate descriptive statistics and perform a confirmatory factor analysis. The respondents' primary occupations included: students (51.9%), IT staff (15.2%), government employees (13.3%), professionals (7.6%), self-employees (4.1%), housewives (3%), and others (4%). Approximately 73.8% of the sample respondents indicated that they had either undergraduate (64%) or graduate school (9.8%) education.

The respondents were well represented in terms of gender and age. About 30% of the sample respondents had not heard about DMB. Of the 70% of respondents who had heard about DMB, the main sources included: TV (26%), newspaper (20%), Internet (15%), friends (6%), and others (33%). About one-fifth (20.1%) of the respondents were utilizing DMB services. Of those respondents, 62.2% were satisfied with their current DMB service whereas 30.3% were only satisfied on a mediocre level. In other words, only 7.5% of the current DMB users were not satisfied with their DMB services. The current users accessed their DMB phones for news and information; leisure and tourism; public relations (marketing); shopping; games; and education. The users believed that the DMB services would impact service industries such as tourism and retail.

The results also indicated that among the sample respondents, the non-users felt that the following major factors would be taken into consideration when choosing DMB services for the future: (1) pricing of DMB cellular handset, (2) video quality, (3) program content, (4) quality [of DMB cellular handset], (5) ease of use, and (6) others [e.g., customer service by the DMB cell phone manufacturer or service provider; brand image and perception]. The aforementioned results from the sample respondents were very astonishingly similar to the 19 interviewees' perceived values.

The independent variables that determine DMB services are: price, usage time, and program content. The dependent variable is DMB service. Table 2 provides a definition of each of these variables. The reliability measure (construct validity) for these constructs was Cronbach's coefficient (alpha). Even though the general rule of thumb for reliability is a value of 0.8 (alpha), values of 0.6 or 0.7 may be considered adequate in some cases (Hair, 1998). Overall, the model provides a valid representation of the data and the constructs are reliable. The reliability test generated Cronbach's coefficient alpha of .7343 for the 12 items. From the analysis, it was concluded that the measure of 12 items was reliable. Coefficient alphas for the four constructs are shown in Table 3.

A series of principal components factor analyses using a VARIMAX rotation were used to assess the unidimensionality in this study. Eigenvalues of at least 1.0 were used to assess the number of factors to extract. The dimensionality of each factor was assessed by examining factor loadings. Factor loadings on construct are shown in Table 4.

Assessing dimensionality involves examining the inter-correlations among the major constructs. A correlation matrix for the constructs is shown in Table 5. The inter-construct correlation coefficients were all positive and significant at less than 0.01 (see Table 5).

The *t* test was used in the quantitative analysis. The price factor of the DMB phone usage is not an issue if the user perceives the DMB program content to be valuable. Table 6 also showed that DMB service was affected by program content (beta=0.358, t-value=5.689). The users associate easy access/connection time to the DMB services with reliability provided by DMB equipment makers or service providers (beta=0.300, t-value=5.104). H_1 and H_2 were supported. ANOVA and Duncan test were used to evaluate hypotheses H_3, H_4, H_5, and H_6.

Table 3. Coefficient alpha for construct

Construct	Variables	Cronbach's alpha
Price	PR1, PR2, PR3	.7970
Access/Usage time	TM1, TM2, TM3	.6218
Program content	CO1, CO2, CO3	.8104
Service	SE1, SE2, SE3	.7081

Table 4. Construct: Factor loadings

Constructs	Loading	Eigenvalue	Communality (%)
PR1	.875		
PR2	.863	2.137	71.223
PR3	.791		
TM1	.792		
TM2	.744	1.710	56.986
TM3	.727		
CO1	.908		
CO2	.907	2.189	72.970
CO3	.737		
SE1	.838		
SE2	.782	1.897	63.219
SE3	.764		

Table 5. Correlation matrix for the constructs

	Price	Usage time	Program content	Service
Price	1.000			
Access/Usage time	.296**	1.000		
Program content	.454**	.497**	1.000	
Service	.266**	.481**	.511**	1.000

** *P<0.01*

Table 6. Analysis of service performance

Independent Variable	Dependent Variable: Service	
	Beta	t-value
Price	0.008	0.141 (sig = .888)
Access/Usage time	0.300	5.104 (sig = .000)
Program contents	0.358	5.689 (sig = .000)
R^2	0.330	
F	42.602	
Sig.	.000	

DMB PHONE PRICE AND RELATED FEES

The users were asked to rate the importance of price issues of the DMB handset and related service fees when selecting a DMB cell phone. These issues include price per program content, price per usage time, and price of the DMB handset. The mean response among the teens was 4.46 (on a scale of 1=unimportant and 5=very important); 4.23 for 20s, and 4.33 for 30s. The mean response among the older generations (40s and older) was 4.0. Table 7b showed that there were significant differences among teens and the other age groups (20s, 30s, 40s, and older). And the 20s and 30s age group perceived the DMB phone price and related fees differently, when compared with teens, and the 40s and older age group.

In an effort to explain this unexpected finding, the authors used analysis of variance (ANOVA) to see if there were any significant differences between the DMB handset price/related fees and age group. As shown in Table 7a, the difference is statistically significant (F=12.583, df=3, 260, p=0.000), which demonstrates that the younger generation is willing to pay the current market price for the DMB handset and related services, given that they perceive the content to be useful and worthwhile. This supports H_3, as it validates that there is a difference between the age groups and their perceptions of DMB phone price.

ACCESS/USAGE TIME

The users were asked to rate the importance of access/usage time issues of DMB services and handset when selecting a DMB cellular phone. These issues include access time, air time, and the time it takes to get familiarized with the DMB handset and services. The mean response among the teens was 4.02 (on a scale of 1=unimportant and 5=very important); 3.82 for 30s, 3.86 for 40s and older. Table 8b demonstrates a slight discrepancy between teens and those in their 20s, but no significant divergences among other age groups (30s, 40s and older). The ANOVA test showed that there was not a significant difference between age groups and their judgments of the importance placed on the DMB access/usage time (see Table 8a). This does not support H_4, as it validates that there is no difference between age groups and their percep-

Table 7. ANOVA and Duncan Test of DMB phone price and related fees

	Sum of squares	df	Mean Square	F	Sig.
Between groups	11.161	3	3.720	12.583	.000
Within groups	76.879	260	.296		
Total	88.040	263			

7a. ANOVA

Age	N	Subset for alpha = .05		
		1	2	3
Teens	45			4.6889
20s	116		4.2328	
30s	52		4.3269	
40s and older	51	4.0261		
Sig.		1.000	.354	1.000

7b. Duncan Test

tions of the DMB phone access/usage time. When focusing on strategic moves, the major players in the DMB market do not have to place as much emphasis on the end-users' access/usage time.

PROGRAM CONTENT

The users were asked to rate the importance of program content of DMB handsets when selecting a DMB cellular phone. These issues include video quality of content, audio quality of content, and a selection of content. As shown in Table 9a, the ANOVA test showed that there was a difference between various program contents and age groups ($F = 6.30$, $df = 3, 260$, $p=0.000$). The mean response among the teens was 4.69 (on a scale of 1=unimportant and 5=very important); 4.29 for the 20s age group, 4.30 for the 30s age group, and 4.32 for the 40s age group and older. Table 9b showed significant deviation between teens and the other age groups (20s, 30s, and 40s and older).

Table 8. ANOVA and Duncan Test of access/usage time

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	2.943	3	.981	2.502	.060
Within groups	101.935	260	.392		
Total	104.878	263			

8a. ANOVA

Age	N	Subset for alpha = .05	
		1	2
Teens	45		4.0296
20s	116	3.7328	
30s	52	3.8269	3.8269
40s and older	51	3.8627	3.8627
Sig.		.298	.102

8b. Duncan Test

Table 9. ANOVA and Duncan of program content

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	5.870	3	1.957	6.304	.000
Within groups	80.689	260	.310		
Total	86.559	263			

9a. ANOVA

Age	N	Subset for alpha = .05	
		1	2
Teens	45		4.6963
20s	116	4.2902	
30s	52	4.3013	
40s and older	51	4.3268	
Sig.		.743	1.000

9b. Duncan Test

The perception on program content for each age group (20s, 30s, and 40s and older) differed from the teens' perceptions. This supports H_5.

SERVICE

The users were asked to rate the importance of DMB services when selecting a DMB cellular phone. These services include after-service of the DMB equipment maker or service provider, performance of the DMB phone device, and the credibility of the DMB equipment maker or service provider. As shown in Table 10a, the ANOVA test showed that there was a difference between the age groups and their approach to the importance of DMB services ($F = 3.355$, $df = 3, 260$, $p < 0.019$). The mean response among the teens was 4.54. Table 10b showed that there are significant differences among teens and other age groups (20s, and 40s and older). And each age group's (20s, and 40s and older) perception on services deviated significantly from the teens' perceptions. This supports H_6.

As shown previously, all hypotheses (except H_4) were supported.

ACTUAL USAGE AND IMPLICATIONS

Although the exploratory study's results showed teens and the 20s age group as heavy users of DMB services, the actual DMB statistical usage data (see Figure 4) differed. The actual usage data was recently released from TU-Media (S-DMB service provider), which revealed that those in their late 20s, 30s, 40s, and 50s represent a large percentage of users of the following DMB services: soap operas, sports, and music program content (Suh, 2005). The results in Figure 5 show a correlation between preferences across age and gender. While the sports channel was the only preferred program among males to spread across all age groups, soap opera programs were preferred by female teens and those in their 20s. The authors believe that there are several reasons as to why the younger consumers may not be currently subscribed to S-DMB services: (1) S-DMB handsets (which retails for $600-$800) are too expensive; (2) The teens lack the extra out-of-pocket money to pay for the S-DMB $13 monthly service fee (and $20 activation fee); and (3) The parents do not feel justified in purchasing a DMB handset for their

Table 10. ANOVA and Duncan Test of Service

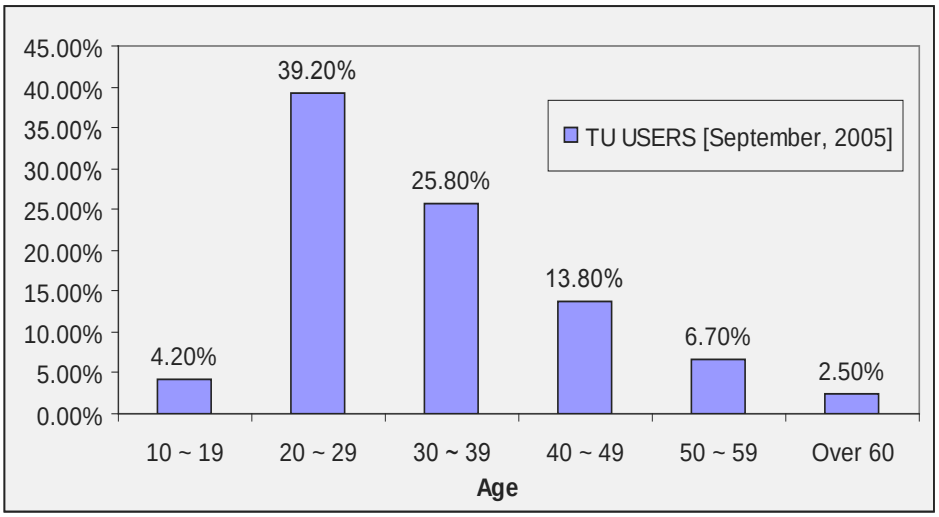
	Sum of Squares	df	Mean Square	F	Sig.
Between groups	3.428	3	1.143	3.355	.019
Within groups	88.557	260	.341		
Total	91.985	263			

10a. ANOVA

Age	N	Subset for alpha = .05	
		1	2
Teens	45		Teens
20s	116	4.2241	20s
30s	52	4.3782	30s
40s and older	51	4.3137	40s and older
Sig.		.185	Sig.

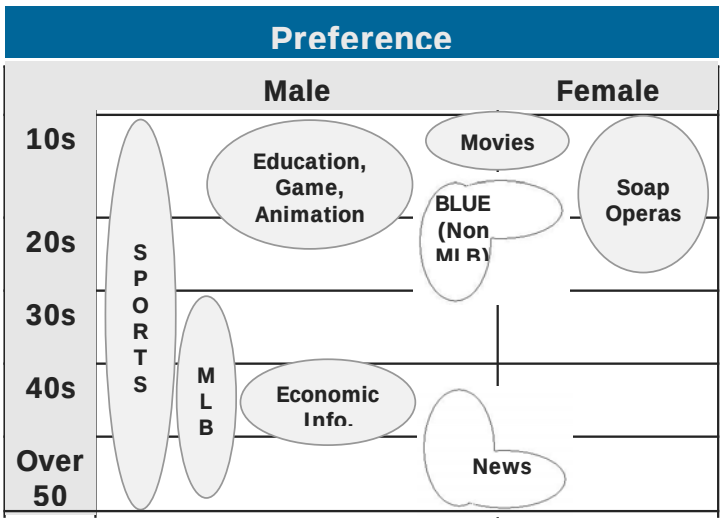
10b. Duncan Test

Figure 4. Actual DMB statistical usage data by age



Source: TU Media. (2005, September)

Figure 5. Preferences for various age groups and genders



Source: TU Media (2006)

children's TV and gaming purposes. Additionally, most of the school-age children have little time to watch DMB program content due to the academic load. Furthermore, the actual usage results among various demographic groups for T-DMB services, once released, are expected to differ from that of S-DMB services given that T-DMB services are free and advertiser-supported.

CONCLUSION

The mobile TV standards (e.g., DMB, DVB-H, and MediaFLO) and wireless technologies will add a new dimension to the connectivity between enterprises and consumers as well as their access to information and entertainment. Given the demand for ubiquitous computing in an impatient,

technology-hungry, instant gratification-seeking population, the desire for mobile TV will continue to grow and soon mobile TV will be synonymous to today's radio in the long run (Kim, 2004). Similar to the interactive TV (iTV) (Tsaih, Chang, & Huang, 2005), the DMB has implications, which include: (1) service and content providers use the DMB as a vehicle for business-to-consumer (B2C) commerce via programs, content, and services; (2) consumers have real-time access to DMB services and programs on mobile phones, PDAs, and other mobile devices anytime and anywhere.

As mentioned earlier, the key issues for the DMB market include: (1) optimal capital investment levels to achieve adequate service coverage for T- DMB and S- DMB services; and (2) appropriate business models, with respect to advertising-supported vs. subscription services. The wireless mobile service industry has very complex issues, which span across technical, logistical, social, and cultural issues (Trappey, Trappey, Hou, & Chen, 2004). Thus, this requires cooperation among the cellular and network service providers, service developers, and equipment makers to collaborate with the government and users to create growth in the cellular telecommunications industry.

Although this research is based on exploratory methods, it still has its limitations. For example, the sample size was collected during the experimental/trial stages of S-DMB services in Korea. The authors reinforce the continuation of this research to solidify findings with an increased sample size of respondents collected during the actual stage of S-DMB and T-DMB services. In addition to this belief, the authors endorse the notion of longitudinal studies conducted to obtain more results. Furthermore, the authors strongly believe that the findings from this exploratory research will be valuable for the DMB service and content providers to gain insight into various age groups and their perceptions.

One of the implications of this paper of wireless mobile technologies and mobile TV is to demonstrate how important the government initiative can benefit less-developed and developing countries. For example, South Korea's Ministry of Information & Communication (MIC) established the "IT839" Strategy—"8" services; "3" infrastructures; "9" growth engines—as a roadmap for Korea's future IT development plan (MIC, 2005). The authors hope that this discussion will be beneficial for the mobile wireless industry and the academia for insight and understanding of the trends of the ubiquitous computing era.

NOTE

* A portion of this chapter is based on an earlier work: Shim, Ahn, & Shim. (2006). Empirical Findings on Perceived Use of Digital Multimedia Broadcasting Mobile Phone Services. *Industrial Management & Data Systems, 106* (2).

REFERENCES

The Australian. (2003, March 4th).

Budde, P. (2002). Asia and Australia telecommunications industry overview. *Annual Review of Communications, 55,* 243-250.

Daft, R., & Lengel, R. (1986). Organizational formation requirements, media richness and structural design. *Management Science, 32*(5), 555-571.

Eom, M. (2006). T-DMB overview in Korea. In *Proceedings of 2006 Wireless Telecommunications Symposium,* Pomona, CA.

Feagin, J. R., Orum, A. M., & Sjoberg, G. (1991). *A case for the case study.* Chapel Hill, NC: The University of North Carolina Press.

Gharavi, H., Love, P., & Cheng, E. (2004). Information and communication technology in the stockbroking industry: An evolutionary approach to the diffusion of industry. *Industrial Management & Data Systems, 104*(9), 756-765.

Gilbert, D., Lee-Kelley, L., & Barton, M. (2003). Technophobia, gender influence and consumer decision-making for technology-related products. *European Journal of Innovation Management, 6*(4), 253-263.

Hair, J. F. (1998). *Multivariate data analysis.* Prentice Hall.

Kim, J. (2004). Terrestrial DMB's effects on the electronics industry. In *Proceedings of 2004 Terrestrial DMB International Forum* (pp. 131-142).

KORA Research. (2004, May). *A market policy study on DMB* (Rep. No. 2003-10).

Korea Radio Station Management Agency. (2004). *A market policy study on DMB.*

Kim Tae-gyu, K. (2005, January 18). Korea's free mobile broadcasting faces snag. *The Korea Times.*

Korean Society for Journalism and Communication Studies. (2003). *A study on satellite DMB.*

Kvale, S. (1983). The qualitative research interview: A phenomenological and a hermeneutical mode of understanding. *Journal of Phenomenological Psychology, 14*(2), 171-196.

Larsen, T., & Sorebo, O. (2005). Impact of personal innovativeness on the use of the Internet among employees at Work. *Journal of Organizational and End User Computing, 17(*2), 43-63.

Lee, S. M. (2003). South Korea: From the land of morning calm to ICT hotbed. *Academy of Management Executive, 17(*2), 7-18.

Ministry of Information and Communication. (2005). *IT 839 Strategy.* Republic of Korea.

Nyberg, A. (2004). Positioning DAB in an increasingly competitive world. In *Proceedings of 2004 Terrestrial DMB International Forum* (pp. 131-142).

Olla, P., & Atkinson, C. (2004). Developing a wireless reference model for interpreting complexity in wireless projects. *Industrial Management & Data Systems, 104*(3), 262-272.

QUALCOMM Incorporated. (2005). *MediaFLO: FLO technology brief.* Retrieved from www.qualcomm.com/mediaflo

Rogers, E. M. (1983). *Diffusion of innovation* (3rd ed.). New York: The Free Press.

Shim, J. P. (2005a). Korea's lead in mobile cellular and DMB phone services. *Communications of the Association for Information Systems, 15,* 555-566.

Shim, J. P. (2005b). Why Japan and Korea are leading in the mobile business industry. *Decision Line, 36*(3), 8-12.

Shim, J. P., Ahn, K. M., & Shim, J. (2006). Empirical findings on the perceived use of digital multimedia broadcasting mobile phone service. *Industrial Management & Data Systems, 106*(2), 155-171.

Shim, J. P., Shin, Y. B., & Nottingham, L. (2002). Retailer Web site influence on customer shopping: An exploratory study on key factors of customer satisfaction. *Journal of the Association for Information Systems, 3,* 53-75.

Shim, J. P., Varshney, U., & Dekleva, S. (2006a). Wireless evolution 2006: Cellular TV, wearable computing, and RFID. *Communications of the Association for Information Systems, 18,* 497-518.

Shim, J. P., Varshney, U., Dekleva, S., & Knoerzer, G. (2006b). Mobile and wireless networks: Services, evolution & issues. *International Journal of Mobile Communications, 4*(4), 405-417.

Smagt, T. (2000). Enhancing virtual teams: Social relations v. communication technology. *Industrial Management & Data Systems, 100*(4), 148-156.

Suh, Y. (2005). *Current overview of S-DMB.* TU Media.

Teng, R. (2005, January). Digital multimedia broadcasting in Korea. *In-Stat Report No. IN-052469WHT.* Retrieved from http://www.cctv.org/InStatPaper.pdf

Thompson, C. J., Locander, W. B., & Pollio, H. R. (1989). Putting consumer experience back into consumer research: The philosophy and method of existential-phenomenology. *Journal of Consumer Research, 16,* 133-146.

Trappey, A., Trappey, C., Hou, J., & Chen, B. (2004). Mobile agent technology and application for online global logistic services. *Industrial Management & Data Systems, 104(*2), 169-183.

Tsaih, R., Chang, H., & Huang, C. (2005). The business concept of utilizing the interactive TV. *Industrial Management & Data Systems*, *105*(5), 613-622.

Ventatesh, W. (2000). Age differences in technology adoption decisions: Implications for a changing work force. *Personnel Psychology, 53,* 375-403.

ENDNOTES

1. http://www.pcworld.com/news/article/0,aid,124478,00.asp (2006, January 24)
2. Wi-Fi Hotstats. *Wireless Review, 22*(8) (August, 2005)
3. www.scala.com/vignettes/digital-multimedia-broadcasting.html
4. www.chiefexecutive.net/depts/technology/197a.htm
5. http://www.economist.com/business/displaystory.cfm?story_id=5356658&no_jw_tran=1&no_na_tran=1 (2006 Jan 5)

This work was previously published in Global Mobile Commerce: Strategies, Implementation, and Case Studies, edited by W. W. Huang, Y. Wang, and J. Day, pp. 234-252, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.15
Motivations and Barriers to the Adoption of 3G Mobile MultiMedia Services:
An End-User Perspective in the Italian Market

Margherita Pagani
I-LAB Centre for Research on the Digital Economy, Bocconi University, Italy

Danilo Schipani
Valdani Vicari & Associati, Italy

ABSTRACT

This chapter provides an end-user perspective on mobile multimedia services that are likely to emerge with the roll out of Third Generation Mobile Services (3G). More specifically, the objectives of the study are:

- To provide an insight into current behavior in terms of attitudes towards, access and usage of multimedia mobile services by current end users;
- To establish main clusters of mobile users;
- To investigate the possible motivations and barriers of usage of new mobile multimedia services as viewed by current users.

The remainder of this chapter is organized into the following four sections. The first section provides a brief review of the literature on the technology acceptance model. Next we present our research model based on a qualitative exploratory survey conducted in six markets. Then we test the proposed model on the Italian market and present the analysis and results of our study. Finally we make conclusions by discussing the implications of our study, followed by presenting future research directions.

INTRODUCTION

As telecommunications move into an era where the distinction between voice, video and data will be

blurred, convergence of communications, information, entertainment, commerce and computing will lay the foundation for the development of an Information Society.

Over the last five years there have been a number of significant developments in multimedia computing power, CD-ROM technology, digital television, the Internet/Intranet, IP-based services, and terrestrial and satellite mobile communications, which could have a profound impact on our society. These technologies and systems may enable dramatic changes to take place in working practices, entertainment, education and health care.

Many organizations within the computing, entertainment, and communications industries are now looking to identify and capitalize on the promise of new market opportunities in multimedia created by these developments.

However, demand for multimedia services, should they be successful, is unlikely to be constrained to the fixed network. Greater pressure on time, and the need for flexibility and responsiveness in business, will lead to a growing demand for access to these services anytime, anywhere.

In order to meet the evolving needs of customers, and to capture the opportunity which this evolution represents, the mobile industry is looking to define and develop a third generation of mobile technology that will take the personal communications user into the Information Society by delivering voice, graphics, video and other broadband information direct to the user, regardless of location, network or terminal.

The purpose of the chapter is to provide an end-user perspective on mobile multimedia services that are likely to emerge with the roll out of Third Generation Mobile Services (3G).

The remainder of this chapter is organized into the following four sections. The first section provides a brief review of the literature on the technology acceptance model. Next we present our research model based on a qualitative exploratory survey conducted in six markets. Then we test the proposed model on the Italian market and present the analysis and results of our study. Finally we make conclusions by discussing the implications of our study, followed by presenting future research directions.

TECHNOLOGY ACCEPTANCE MODEL (TAM): THE THEORETICAL BACKGROUND

Information Systems (IS) researchers have made significant efforts in building theories to examine and predict the determinant factors of information technology (IT) acceptance (Agarwal & Prasad, 1998, 1999). Existing models of IT acceptance have their foundations from several diverse theories, most noticeably innovation diffusion theory, where individuals' perceptions about using an innovation are considered to affect their adoption behaviors (Agarwal & Prasad, 1998; Moore & Benbasat, 1991; Rogers, 1995). Other important theoretical models that attempt to explain the relationship between user beliefs, attitudes, intentions, and actual system use include the theory of reasoned action (TRA) (Ajzen & Fishbein, 1980), the theory of planned behavior (TPB) (Ajzen, 1991), and the technology acceptance model (TAM) (Davis, 1989; Davis et al., 1989). Although there are numerous studies in the field of adoption and diffusion of marketing-enabling technology (Daghfous, Petrof & Pons, 1999; Holak & Lehman, 1990; Labay & Kinnear, 1981; Plouffe, Vandenbosch & Hulland, 2001; Rogers, 1995), previous work has mainly focused on the adoption of products and technology (Au & Enderwick, 2000; Davis, 1989; Eastlick & Lotz, 1999; Verhoef & Langerak, 2001). In contrast, the perspective on services and service-enabling technologies is considerably less pronounced. Despite the fact that several trend studies have been conducted regarding the potential of wireless technology and 3G services (Durlacher, 2001; UMTS Forum, 2001), there

exists a need for more substantive, theory-based research, creating a more in-depth understanding of consumer behavior with regard to m-commerce. In the information system literature on IT adoption, researchers have conducted several studies to examine the relationship between perceived ease of use, perceived usefulness, and the usage of other information technologies (Adams et al., 1992; Chau, 1997; Davis, 1989; Davis et al., 1989, Hendrickson & Collins, 1996; Mathieson, 1991; Szajna, 1996). Their researches have supported the Technology Acceptance Model (TAM) proposed by Davis (1989), which posits that perceived ease of use and perceived usefulness can predict the usage of technology.

TAM was derived from the Theory of Reasoned Action (TRA). According to Davis (1989), perceived usefulness and perceived ease of use are the two determinants that influence people's attitude toward IT usage intention and actual IT usage. Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" and perceived ease of use is defined as "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320). Davis and his colleagues (Davis, 1989; Davis et al., 1989, 1992) demonstrated that perceived ease of use affected usage intention indirectly via perceived usefulness.

In an extension to TAM, Davis and his colleagues examined the impact of enjoyment on usage intention (Davis et al., 1992). They reported two studies concerning the relative effects of usefulness and enjoyment on intention to use and usage of computers. As expected, they found enjoyment had a significant effect on intention. A positive interaction between usefulness and enjoyment was also observed.

Several recent empirical studies have validated adoption theory in relation to a wide range of products (Holak & Lehman, 1990; Labay & Kinnear, 1981; Ostlund, 1973; Rogers, 1995) and technology (Beatty, Shim & Jones, 2001; Plouffe et al., 2001). A large number of studies have investigated the use of electronic commerce, but the field of mobile commerce has been left virtually unexplored. In this research our goal is to extend the TAM model to study motivations and barriers to the adoption of 3G mobile multimedia services. In the following sections the research is divided into two stages: an exploratory qualitative stage followed by a quantitative stage focused on the Italian market.

RESEARCH FRAMEWORK

Methodology

Many factors positively or negatively influence users' adoption of multimedia mobile services. In this section we identify several variables that influence adoption of 3G mobile multimedia services. The variables are derived from two preliminary pilot studies realized on a sample of young people in Italy and USA followed by an exploratory qualitative study conducted by Nokia through 24 focus groups in six markets (Brazil, Germany, Italy, Singapore, UH, USA).

The second stage of the analysis concentrates specifically on a quantitative marketing research. Data were gathered by means of a questionnaire. The population consists of 1,000 Italian users of mobile services. It tries to describe behaviors, roles and test variables influencing adoption of mobile computing. We consider Italy because it is the European country with the higher penetration of mobile phones and profitability, and it is also prone to market innovation.

The main goal of this research is to identify a hierarchy of importance concerning the critical factors influencing the adoption of mobile services. To realize this research objective, conjoint analysis was seen as the appropriate statistical tool.

Table 1. Fieldwork details

Country	Sample	Field times
Brazil	Nationally representative o f adults aged 1 8-64 w ho a re e conomically active	6th – 20th March 2001
Germany	Nationally r epresentative of adults aged 14+	23rd March – 5th April 2001
Italy	Nationally r epresentative of adults aged 15+	23rd March – 5th April 2001
Singapore	Nationally representative o f adults aged 15-64	13th – 26th April 2001
UK	Nationally representative o f adults aged 15+	23rd March – 5th April 2001
USA	Nationally r epresentative of adults aged 18+	21st – 30th March 2001

Conjoint Analysis

Conjoint analysis is a technique that allows a set of overall responses to factorially designed stimuli to be decomposed so that the utility of each stimulus attribute can be inferred from the respondent's overall evaluations of the stimuli (Green, Helsen & Shandler, 1988). A number of (hypothetical) combinations of service elements can be formulated that will be presented to a sample of customers. According to Lilien and Rangaswamy (1997), the analysis comprises three stages.

The first stage is concerned with the design of the study, where the attributes and levels relevant to the product or service category will be selected. In the second stage customers rate the attractiveness of a number of possible combinations of customer service elements. Finally, in the third stage, ratings are used to estimate part-worth utilities, that is, the utility that is attached to the individual levels of each service element included in the research design. Consequently, an accurate estimate of customer trade-offs between services elements can be obtained.

The dependent variable in our study was the intention to make use of mobile services.

Exploratory Qualitative Stage

The fieldwork has been carried out face to face in the first and second quarters of 2001 through 24 focus groups conducted by Nokia Networks in six markets (Brazil, Germany, Italy, Singapore, UK, USA). The interviews focused in on the core target for the 3G offering, namely, teenagers, young adults and family adults, all currently using mobile phones for personal usage. The sample was segmented by age, 16-19, 20-29 and 30-45, and by life stage.

The research looked primarily at the following mobile multimedia services: photo messaging, mobile e-mail, video messaging and postcard messaging. However, the research also briefly touched on rich text messaging, and on video calling.

Of utmost importance in the study was to ensure that the respondents concentrated on the messaging format, and did not allow previous misconceptions about service or delivery of the service. They were therefore told to imagine that there would be no network problems, and not to concentrate on pricing.

The prompted statements offered to the sample as motivations for usage of the future multimedia mobile services can be classified to form eight broad segments of usage (Table 2):

Table 2. Motivation segmentation

1. Business
for business purposes
2. Formality
When I want to send a formal message
3. Urgency
When I need to know the message has arrived
When I want to send urgent communication
As a rapid way to stay in touch
4. Function
To send a long piece of text
To send an attachment
When I don't feel like talking
Practical reason (like to show something I want to buy)
5. Price
When I want to communicate cheaply
6. Discretion
Need to be discreet and quiet
When talking would disturb people around me
Might disturb the person I'm trying to contact
7. Personal contact
To keep in touch with friends/family abroad
To send an intimate message
To contact people I don't see very often
As a personalized way to send a message
To increase the feeling of contact
To share an experience
Nice for people to see me if they haven't done so for a while
For longer greetings
When I don't want to talk, but need to communicate
8. Fun
Joke or chit-chat with friends
As a novel way to message
To share an experience
As it is just great fun
To send pictures from my holiday
To show something like a view
To express creativity

1. Business
2. Formality
3. Urgency
4. Function
5. Price
6. Discretion
7. Personal Contact
8. Fun

The research model to be empirically tested in the Italian market is illustrated in Figure 1. The model is derived from the theories and hypothesis described in the preceding section. The relationship constituting the model also has support from prior theoretical and empirical work in the exploratory qualitative stage.

Exploratory Quantitative Stage

A following stage of analysis concentrates specifically on a quantitative marketing research conducted in the second quarter of 2002 through questionnaires on a sample of 1,000 Italian users of mobile (sampled among over 18 Italians).

One thousand interviews provide a sampling error (at 50%) of 3.1% (with a probability level of 95%).

The research, managed through telephone calls, tries to describe behaviors, roles and variables influencing adoption of mobile computing.

The results of the quantitative marketing research are now summarized. This research was structured in order to deepen the motivations and barriers towards the innovative services delivered through 3G mobile services, the eventual levels of demand and usage and the content types and formats that consumers express opinion for.

Key items in the questionnaire used for analyzing the survey are as follows:

1. **Degree of service innovation** perceived by consumers. Respondents selected their answers from a list of innovative services categories;
2. **Interest** for the services categories under scrutiny;
3. **Preference** for means/platforms through which selected services can be accessed (portables, phone and/or TV);
4. **Analysis** of key features of services (ease of use, speed, cost and usefulness);
5. **Ranking** of services features.

Figure 1. Adapted TAM model on the adoption of multimedia mobile services

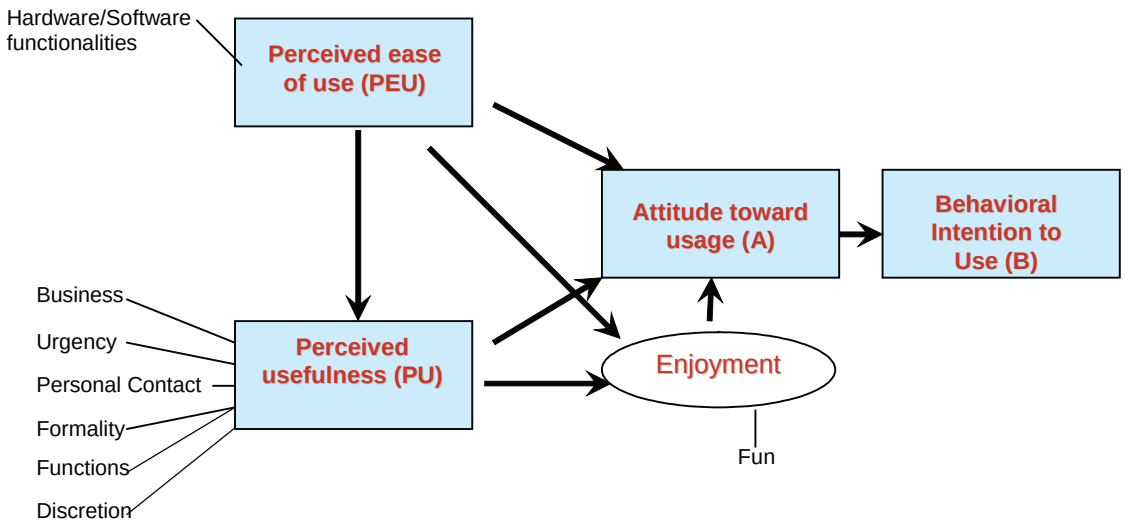
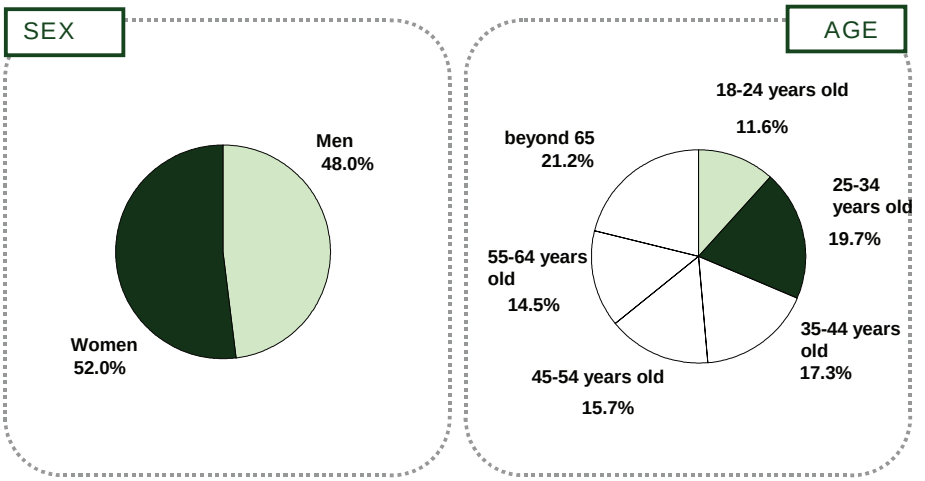


Figure 2. Composition of the sample



The services considered in the questionnaire are the following:

- Interactive and real-time entertainment;
- Data exchange among people and between people and various electronic devices;
- Contextual and real-time shopping;
- Portfolio and personal funds management;
- Safety-related services;
- Location-based services.

All the services have been considered rather innovative (the average is 7.1 on 1 minimum -9 maximum scale).

In terms of the interest expressed towards these services, the sample distributes are shown (see Figure 3).

Table 3 shows the main features preferred by the people to be attracted to use these services.

"Usefulness" and ease of use are considered the most important variables in order to access the segments of population and, as shown in Figures

Figure 3. Interest expressed towards multimedia mobile services

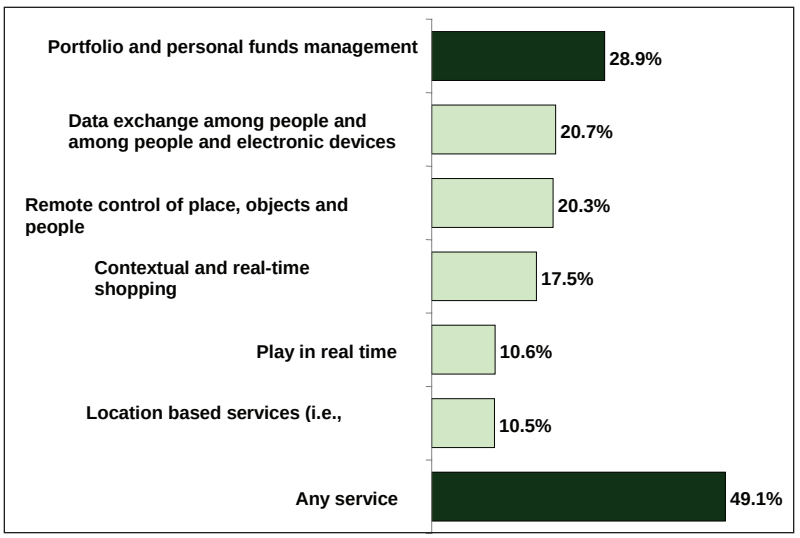


Table 3. Features preferred

	IMPORTANCE	
	Ranking	%
Usefulness	**1°**	31.3%
Ease of use	**2°**	26.7%
Price	**3°**	23.8%
Speed of use	**4°**	18.2%

4 and 5, there are different meanings assigned to these words.

The final objective of the research was to identify the key descriptive elements of homogeneous segments of the population. This is relevant in order to define the right strategies to offer the new services in the proper and differentiated way.

The most statistically powerful variable in order to distinguish the behaviors of people is the degree of interest towards the innovative services.

If we then clusterize the sample using this variable, and cross it with the socio-demo data, it turns out that the kind of activities performed in life by the consumers is the strong predictor of their future use of the new services.

In particular, it is possible to describe two different segments as indicated in the following figure:

- Cluster 1 is composed of people who declared they are not interested in the new services;
- The remaining 51% can be divided in two groups that are different in terms of the way firms should approach them to sell the new services.

The two segments are:

- The "professionals," that is, people who mainly are managers or entrepreneurs in life, who are 38% of the interviewed base;
- The "students," who account for the remaining 13%.

The purpose is now to identify the variables network operators can use to access the identified clusters. This is an essential piece of information for crafting the right strategies in order to "catch" the segments.

The "professional" segment is made of people who look for *usefulness* as the almost exclusive

Figure 4. Meaning of usefulness

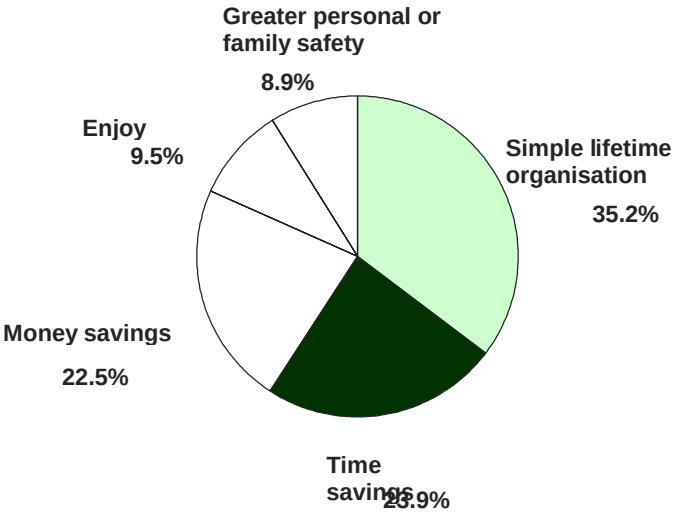
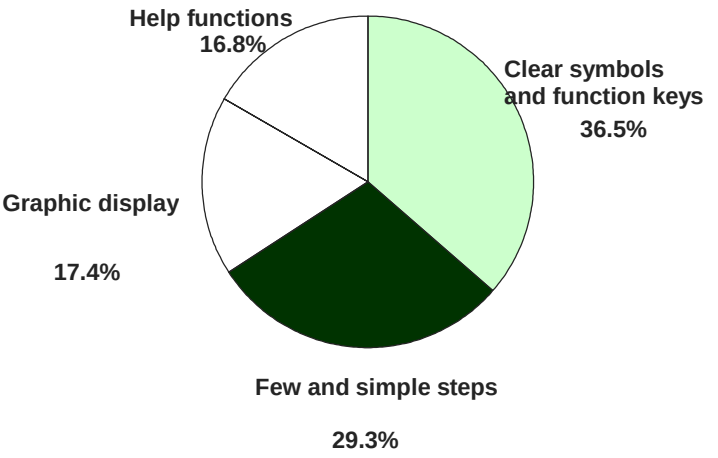


Figure 5. Meaning of ease of use



variable in order to access and pay for the service.

The "students" segment is made of people who look mainly for *low-cost* and *convenience*.

For the entire interviewed base, an interesting relationship emerges: the degree of interest is inversely related to the degree of knowledge of the service. In particular it has been noticed that people who declare a low level of interest in these services are those who actually know least the main features and potential outcomes of these services, even though the interviewer deeply explained the meaning of each service.

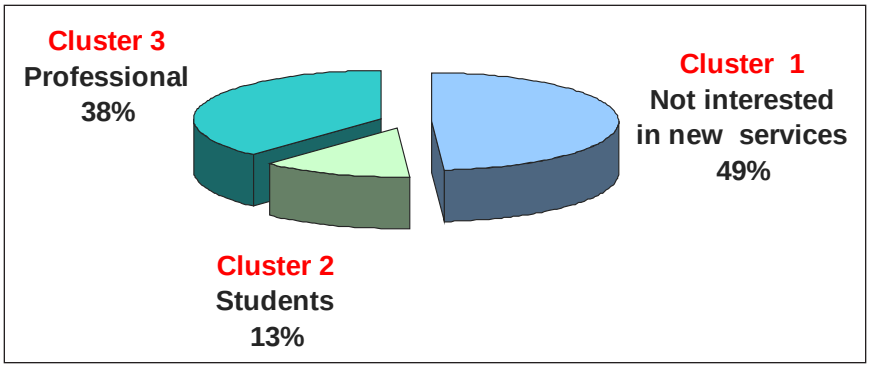
CONCLUSION

In this research, we attempt to identify valid factors that predict a user's adoption of 3G mobile multimedia services.

The findings show key characteristics and factors playing decisive roles in the development of strategies for the launch of multimedia mobile services.

The findings of this study have significant implications also in the perspective of research on mobile consumer behavior. Our study provides further evidence on the appropriateness of using the TAM model to measure the different dimensions of actual multimedia mobile usage and it

Figure 6. Main clusters of mobile users in the Italian market (base 1,000 Italian mobile users)



provides empirical evidence that PEU (perceived ease of use) and PU (perceived usefulness) are important factors that influence the user's adoption of 3G multimedia mobile services.

The findings of the study suggest important practical implications for businesses currently providing mobile multimedia services as well as those that are planning to do so. It is evident from this study that in order to influence adoption of 3G multimedia services, perceived ease of use (PEU) and perceived usefulness (PU) must be enhanced.

ACKNOWLEDGMENT

The authors wish to acknowledge Massimo Farioli for his cooperation in the data processing phase.

REFERENCES

Adams, D.A., Nelson, R.R.,& Todd, P.A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly, 16*(2), 227-250.

Agarwal, R., & Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences, 28*(3), 557-581.

Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research, 9*(2), 204-215.

Agarwal, R., & Prasad, J. (1999). Are individual differences germane to the acceptance of new information technologies? *Decision Sciences, 30*(2), 361-391.

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes, 50*(2), 179-211.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour.* Eaglewood Cliffs, NJ: Prentice-Hall.

Allwood, C.M. (1998) *Human-computer interaction-a psychological perspective*. Lund, Sweden: Studentlitteratur.

Allwood, C.M., & Ljung, K. (1999). Computer consultants' view of user participation in the system development process. *Computer in Human Behavior, 15*, 713-734.

Briggs, R.O., Adkins, M., Mittleman, D., Kruse, J., Miller, S., & Nunamaker, J.F., Jr. (1998). A technology transition model derived from field investigation of GSS use aboard the U.S.S. CORONADO. *Journal of Management Information Systems, 15*(3), 151-195.

Briggs, R.O., Vreede, G.J. de, & Nunamaker, J.F., Jr. (2003). Collaboration engineering with ThinkLets to pursue sustained success with group support systems. *Journal of Management Information Systems, 19*(4), 31-63.

Chau, P.Y.K. (1997). Re-examining a model for evaluating information center success using a structural equation modelling approach. *Decision Sciences, 28*(2), 309-334.

Chin, W.W., & Todd, P.A. (1995). On the use, usefulness, and ease of use of structural equation modelling in MIS research: A note of caution. *MIS Quarterly, 19*(2), 237-246.

Colby, C. (2002). Techno-ready marketing of e-services: Customer beliefs about technology and the implications for marketing e-services. In T.R. Rust & P.K. Kannan (Eds.), E-service: New directions in theory and practice. Armonk, NY: M.E. Sharpe Inc.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1003.

Doll, W.J., & Torkzadeh, G. (1998, June). The measurement of end-user computing satisfaction. *MIS Quarterly, 12*(2), 259-274.

Gefen, D., & Straub, D.W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly, 21*(4), 389-400.

Mathieson, K. (1991). Predicting user intensions: Comparing the technology acceptance model with theory planned behaviour. *Information Systems Research, 2*(3), 192-222.

Moore, G., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192-222.

Nokia. (2002) *3G market research mobile messaging: An end user perspective*. Nokia Report.

Parasuraman, A (2000). Technology readiness index: A multiple item scale to measure readiness to embrace new technologies. *Journal of Service Research, 2*(4), 307-320.

Rogers, E. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Straub, D., Limayem, M.,& Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science, 41*(8), 1328-1342.

Szajna, B. (1994). Software evaluation and choice: Predictive validation of the technology acceptance instrument. *MIS Quarterly, 18*(3), 319-324.

Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. *Management Science, 42*(1), 85-92.

Taylor, S., & Todd, P.A. (1995). Understanding information technology usage; A test of competing models. *Information Systems Research, 6*, 144-176.

VVA. (2002). *Osservatorio marche: Le telecomunicazioni*. VVA Report.

Zhu, W., & Nah, F.-H. (2002). Factors influencing adoption of mobile computing. *Issues and Trends of IT Management in Contemporary Organizations – IRMA Conference Proceedings.*

This work was previously published in E-Commerce and M-Commerce Technologies, edited by P. C. Deans, pp. 80-95, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 1.16
A Proposed Framework for Mobile Services Adoption:
A Review of Existing Theories, Extensions, and Future Research Directions

Indrit Troshani
University of Adelaide, Australia

Sally Rao Hill
University of Adelaide, Australia

ABSTRACT

Mobile services are touted to create a significant spectrum of business opportunities. Acceptance of these services by users is, therefore, of paramount importance. Consequently, a deeper insight is required to better understand the underlying motivations leading users to adopting mobile services. Further, enhanced understanding would also help designing service improvements and appropriate adoption strategies. Most of the existing theoretical acceptance models available originate from organisational contexts. As mobile services bring additional functional dimensions, such as hedonic or experiential aspects, using extant models for predicting mobile services acceptance by individuals may be inadequate. The aim of this chapter is to explore and critically assess the use of existing acceptance theories in the light of evolving mobile services. Constructs affecting adoption behaviour are discussed and relevant extensions are made which culminate with a framework for mobile services adoption. Managerial implications are explored and future research directions are also identified.

INTRODUCTION

Mobile technologies and services are touted to create a significant spectrum of business opportunities. According to the International Telecommunications Union (ITU) mobile phone penetration rates have increased significantly in many countries in Northern Europe (e.g., Sweden—98.05%, Denmark—88.72%, Norway—90.89%) (Knutsen,

Constantiou, & Damsgaard, 2005). Similarly, Japan and Korea have consistently experienced very high diffusion rates of mobile devices and services (Carlsson, Hyvonen, Repo, & Walden, 2005; Funk, 2005). While experts predict that by 2010 online access via mobile channels is expected to reach 24% of homes in North America, 27% in Eastern Europe, and 33% in North-Western Europe (Hammond, 2001), the current penetration rate in many countries in the Western hemisphere and Asia-Pacific, including the U.S. and Australia lags behind the forerunners (Funk, 2005; Ishii, 2004; Massey, Khatri, & Ramesh, 2005). Given the difference between rapid growth rates in the adoption of mobile technologies and associated services in some countries and the relatively slow growth rates in others (Bina & Giaglis, 2005; Knutsen et al., 2005), it is important to identify the factors and predictors of further adoption and integrate them into a consolidated framework.

Mobile technology is enabled by the collective use of various communication infrastructure technologies and portable battery-powered devices. Examples of mobile devices include notebook computers, personal digital assistants (PDAs) and PocketPCs, mobile, "smart" and Web-enabled phones, and global positioning system (GPS) devices (Elliot & Phillips, 2004). There is a variety of communication infrastructure technologies that can enable these devices. Data networking technologies, such as GSM, GPRS, and 3G, are typically used for connecting mobile phones. WiFi (wireless fidelity) is used for connecting devices in a local area network (LAN). Mobile devices can be connected wirelessly to peripherals such as printers and headsets via the Bluetooth technology and virtual private networks (VPNs) enable secure access to private networks (Elliot & Phillips, 2004). Mobile devices are powered by mobile applications which deliver various services while enhancing flexibility, mobility, and efficiency for users within business and life domains. Despite the availability of technologically advanced mobile devices there is evidence that advanced mobile services which run on these have not been widely adopted (Carlsson et al., 2005; Khalifa & Cheng, 2002). The adoption of advanced mobile services is important for the mobile telecommunications industry because mobile services associated with technologically advanced devices constitute a massive source of potential revenue growth (Alahuhta, Ahola, & Hakala, 2005; Massey et al., 2005).

The adoption of advanced mobile technologies and services requires further research as most of the current technology acceptance models are based on research conducted in organisational contexts (Carlsson et al., 2005), and there has been only limited research from consumers' perspective (Lee, McGoldrick, Keeling, & Doherty, 2003). The features of mobile technologies and services, such as short message service (SMS), multimedia messaging service (MMS), e-mail, map, and location services, allow for single wireless devices, such as mobile phones, to be used seamlessly and pervasively across traditionally distinct spheres of life, such as work, home, or leisure, and with various levels of time commitment and self-ascribed roles (Dholakia & Dholakia, 2004). The interactions of these aspects are more intense than ever before (Knutsen et al., 2005). As mobile technologies and services add other functional dimensions, such as hedonic and/or experiential aspects (Kleijen, Wetzels, & de Ruyter, 2004; Mathwick, Malhotra, & Rigdon, 2001), applying extant theories outright to determine the acceptance and adoption by individual users may be questionable and inadequate (Knutsen et al., 2005).

Moreover, more research is called for in the adoption of mobile technologies because of the levels of complexity and diversity that may be encountered during their adoption. A number of factors contribute to this level of complexity and diversity. First, there is a strong relationship between the mobile devices and their users because the former always carries the identity of the latter (Chae & Kim, 2003). As a result,

spatial positioning and identification of users is easier in the mobile context than in the traditional innovation adoption (Figge, 2004). Second, most mobile devices have limited available resources including memory, processing power, and user interface, which have the potential to offset ubiquity benefits (Chae & Kim, 2003; Figge, 2004). Third, the lifecycle of mobile technologies is usually short, which increases adoption risks because new technologies become rapidly obsolete and may, therefore, need to be replaced by newer ones. During this process, a certain amount of consumer learning might be required before adopters can be confident and satisfied in using the mobile devices and services (Saaksjarvi, 2003). Again, this supports the argument that current models of technology acceptance may not be applied directly in predicting mobile adoption behaviour because they do not reflect the levels of complexity and diversity in the adoption of mobile technologies.

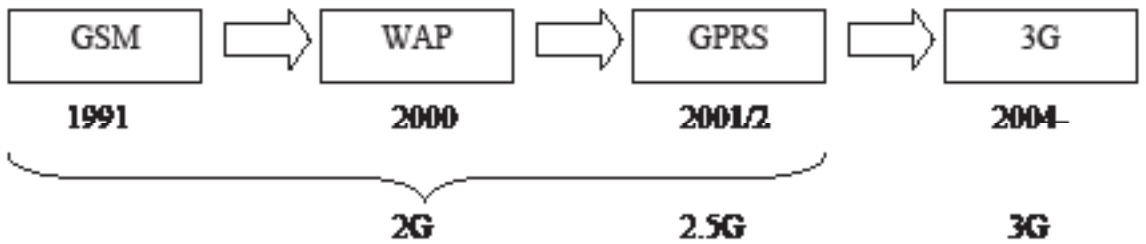
This chapter focuses on mobile phones and the associated services. Examples of mobile services include mobile e-mail, commercial SMS, and MMS services, downloads to portable devices, access to news through a mobile phone, mobile ticket reservations, mobile stock trading, as well other customised services which may be made available by mobile phone operators (Bina & Giaglis, 2005). Research shows that ownership of technologically advanced mobile phones is a main driver for advanced mobile services (Carlsson et al., 2005). Therefore, the adoption of mobile services should also be considered in the context and the technologies which enable them.

The aim of this chapter is to extend the existing models and to propose an integrated conceptual and parsimonious framework which explains adoption behaviour of users of mobile technologies and services. To accomplish this, we first provide an overview of recent developments of mobile technologies and services. Then, a critical assessment of existing acceptance models is made. Next, acceptance constructs and their relevance to mobile technologies and services are discussed. These constructs are then integrated into a new framework about mobile services adoption. In the last section, the implications of this model and future research directions are also discussed.

OVERVIEW OF MOBILE TECHNOLOGY EVOLUTION

In this section, an overview of the evolution of mobile phone technologies is provided. The recognition the evolution of these technologies is important because it puts the adoption constructs discussed later in the appropriate context. The diagram in Figure 1 summarises the evolution of the technologies.

Figure 1. Evolution of wireless technologies (Source: Carlsson et al., 2005)



Second Generation Wireless Devices

The second generation of wireless devices (2G) introduced the digitisation of mobile communication and encompasses several standards which incrementally introduced new services and improved existing ones. It was a big leap forward from the first generation wireless communication (1G) which used analog standards and was characterised by poor quality and narrow bandwidth which resulted in limited adoption by both businesses and individuals (Elliot & Phillips, 2004). The commonly used standards by 2G are the Global System for Mobile Communications, the Wireless Applications Protocol, and the General Packet Radio Service. These are explained in more detail in the following sections.

Global System for Mobile Communication

Launched in the early 1990s, the Global System for Mobile Communications (GSM) constitutes the world's fastest growing and most popular mobile telephony. Available through over 500 networks and serving almost a billion customers in 195 countries, GSM has been expanding exponentially (UMTS, 2003). Most countries, including underdeveloped and developing or even countries with a very low population density have at least two GSM network operators (Rossotto, Kerf, & Rohlfs, 2000). This has increased product and service offerings and competitiveness which has boosted GSM popularity even further.

One of the key advantages of the GSM technology is that it unified a range of different mobile communication standards into a single standard which constitutes a complete and open network architecture. This allows GSM-compatible mobile devices to be connected to any GSM network, therefore, enhancing interoperability. Further, GSM uses digital encoding which encrypts communications between a mobile phone and its base station, which makes interception more difficult. This results in improved security (Elliot & Phillips, 2004).

Another feature of the GSM technology is the automatic country-to-country communication, also known as global roaming. Because international travel for both business and pleasure has increased in recent years, roaming between mobile networks has become valuable as it generates as much as 15% of mobile operator's average revenue per user (ARPU) (UMTS, 2003). The subscriber identity module (SIM) card is another aspect of the GSM technology which is central to its popularity. The SIM card allows operators to manage information about their customers, including customer profile and billing, security access and authentication, virus intrusion and downloading capabilities (UMTS, 2003).

In addition to features such as caller ID, call forwarding, and call waiting, SMS emerged as the first unique mobile service and became the most popular mobile service after 1995 when adopters began using mobiles to send and receive limited amount of data in the form of short messages (Carlsson et al., 2005). While SMS later became the foundational platform for a variety of other services, it is considered to be cumbersome by many users because in addition to memorise service codes, users are also required to type text using the keypad of the mobile device (Carlsson et al., 2005).

Wireless Application Protocol

The Wireless Application Protocol (WAP) was introduced with the aim of providing advanced telephony and data access from the Internet using mobile terminals such as mobile phones, PDAs, smart phone, and other portable handheld devices (van Steenderen, 2002). With WAP, mobile devices can access Web sites specifically designed and built for them. WAP was, therefore, expected to provide the opportunity for connecting two of the fastest growing sectors of the telecommunications industry, namely, the Internet and mobile

communications. As a matter of fact, the hype generated by WAP reached such dizzying heights that from January to August 2000, the number of WAP-compatible Web pages increased from almost zero to 4.4 million (Teo & Pok, 2003).

The benefits of WAP were supposed to extend to several industries ranging from mobile operators, developers of WAP applications, manufacturers of mobile devices, and consumers in terms of various services, including banking, ticket reservations, entertainments, voice and fax mail notifications (Klasen, 2002; van Steenderen, 2002). Nevertheless, except for NTT DoCoMo's i-mode successful demonstration of mobile Internet, WAP has turned out to be a major disappointment with early adopters and other enthusiasts experiencing cognitive dissonance due to the relative oversell (Carlsson et al., 2005; Ratliff, 2002; Teo & Pok, 2003; Xylomenos & Polyzos, 2001).

Other challenges have also adversely affected widespread diffusion of WAP. Narrow bandwidth, low storage memory, and small screen limitations have resulted in slow communications; abridged Internet access has resulted in mediocre interfaces and almost no graphics. This has considerably limited Web site effectiveness (Klasen, 2002). By the end of 2000, only 12 million Europeans had WAP-compatible mobile devices, and of these, only 6% regularly used WAP functionality (Robins, 2003). Worldwide, only 10-15% of whom own WAP-compatible mobile devices would ever use WAP services, suggesting that "WAP had no future" (Klasen, 2002, p. 196). Nonetheless, the introduction of WAP constitutes a major step forward as it showed that Internet browsing is possible in mobile devices in general and phones in particular (Carlsson et al., 2005).

General Packet Radio Service

Simply known as the GPRS, the General Packet Radio Service constitutes an improvement over the GSM technology. GPRS uses packet-based data transfer mechanisms to provide continuous Internet accessibility (Elliot & Phillips, 2004; Hart & Hannan, 2004). With GPRS, users are not required to stay connected all the time in order to use a service. As a result, they are not charged on the basis of the connection time, rather, on the basis of the amount of downloaded data (Carlsson et al., 2005). Overall, GPRS is more efficient and cheaper than GSM, and yet, less widespread among users. Advancements associated with GPRS include the introduction of cameras, colour screens, multimedia messaging service (MMS), and video streaming (Carlsson et al., 2005). Because GPRS enhances 2G services, it is often referred to as the 2.5G technology (Elliot & Phillips, 2004).

Third Generation Wireless Devices

3G represents the next generation of mobile communication technologies, and it makes considerable improvements over its predecessors. These improvements include broad bandwidth which results in higher connection speeds, variety of multimedia capabilities and improved screen display, enhanced security features, and increased storage capacity (Elliot & Phillips, 2004). These enhancements enable users to receive digital photographs, moving video images, high quality sound in their mobile devices, and full unabridged e-mail and Internet access (Elliot & Phillips, 2004). Corporate users are also able to connect remotely to office computers and networks in order to access and download files quickly and easily (Robins, 2003). Because 3G technology mainly improves and enhances many existing services, it is considered to be an evolution rather than a revolution over the previous generation (Carlsson et al., 2005).

3G ensures that anybody, anywhere can access the same services (Grundström & Wilkinson, 2004). Further, 3G aims at integrating both the business and the social domains of the user's life which is the reason why 3G terminals are also referred to as "lifestyle portals" (Elliot & Phillips,

2004, p. 7). Another feature of the 3G technology is its capability to provide location-based services (LBS) which could support health, transport, entertainment, data mining, and so forth (Casal, Burgelman, & Bohlin, 2004; UMTS, 2003). There is evidence that there is demand for such services which constitutes the main economic incentive for the development of the 3G technology (Alahuhta et al., 2005; Repo, Hyvonen, Pantzar, & Timonen, 2004).

A CRITICAL REVIEW OF THEORETICAL MODELS OF TECHNOLOGY ACCEPTANCE

A review of technology acceptance literature revealed many competing theoretical models, each with different focus and tested in different contexts. A significant amount of research effort has been put into building theories to examine how and why individuals adopt new information technologies and predict their level of adoption and acceptance. While one stream of research focuses on individual acceptance of technology (Compeau & Higgins, 1995; Davis, Bagozzi, & Warshaw, 1989), other streams have focused on implementation success at the organizational level (Leonard-Barton & Deschamps, 1988).

Many of the previously empirically researched models have been drawn from social psychology, for example, theory of reasoned action (TRA), motivational model, theory of planned behaviour (TPB), and sociology, for example, social cognitive theory (SCT) and innovation diffusion theory (IDT). Others specifically apply to technology adoption, for example, technology acceptance model (TAM). While each of these models made unique contributions to the literature on technology acceptance and adoption, most of these theoretical models theorise behaviour intention and/or usage as the key dependent variable in explaining acceptance of information technology because behavioural intentions are motivational factors that capture how hard people are willing to try to perform a behaviour (Ajzen, 1991). For example, TPB suggests that behavioural intention is the most influential predictor of behaviour; after all, a person does what s/he intends to do. In a meta-analysis of 87 studies, an average correlation of 0.53 was reported between intentions and behaviour (Sheppard, Hartwick, & Warshaw, 1988). As mobile services and underlying technologies are emerging information technologies, it is appropriate to consider this as the point of departure and use it to form the basis of a theoretical framework in mobile services and technology acceptance and adoption. The models that have been most frequently quoted in the technology acceptance and adoption literature are discussed next.

Theory of Reasoned Action (TRA)

Theory of reasoned action models are considered to be the most systematic and extensively applied approaches to attitude and behaviour research. According to TRA, the proximal determinant of a behaviour is a behavioural intention, which, in turn, is determined by attitude. These models propose that an individual's actual behaviour is determined by the person's intention to perform the behaviour, and this intention is influenced jointly by the individual's attitude and subjective norm. Attitude is defined as "a learned predisposition to respond in a consistently favourable or unfavourable manner with respect to a given object" (Fishbein & Ajzen, 1975, p. 6). A person's attitude towards a behaviour is largely determined by salient beliefs about the consequences of that behaviour and the evaluation of the desirability of the consequences (Fishbein & Ajzen, 1975). Subjective norm is defined as "the person's perception that most people who are important to him think he should or should not perform the behaviour in question" (Dillon & Morris, 1996). In brief, TRA asserts that attitude and subjective norm and their relative weights directly influence behavioural intention.

Theory of Planned Behaviour (TPB) and Decomposed Theory of Planned Behaviour

TPB, which generalizes TRA by adding a third construct—perceived behavioural control (Ajzen, 1991)—has been one of the most influential theories in explaining and predicting behaviour, and it has been shown to predict a wide range of behaviours (Sheppard et al., 1988). TPB asserts that the actual behaviour is determined directly both by behavioural intention and perceived behavioural control. Behavioural intention is formed by one's attitude, subjective norm, and perceived behavioural control (Ajzen, 1991). Further, a decomposed TPB includes constructs such as relative advantage, compatibility, influence of significant others, and risk from the innovation diffusion literature, and decomposing the three perceptions in TPB into a variety of specific belief dimensions. This model offers several advantages over TPB and is considered more complete and management-relevant by focusing on specific factors that may influence adoption and usage (Teo & Pok, 2003).

Technology Acceptance Model (TAM)

TAM can be seen as an adaptation of the theory of reasoned action (TRA) and was developed to predict and explain individual system use in the workplace (Davis, 1989). This model further suggests that two beliefs—perceived usefulness and perceived ease of use—are instrumental in explaining the user's intentions of using a system. Perceived usefulness refers to the degree to which "a person believes that use of the system will enhance his or her performance" whereas perceived ease of use is the degree to which "a person believes that using the system will be free of effort". Simply put, a technology that is easy to use and is useful will lead to a positive attitude and intention towards using the technology.

The main advantage of this model over others is that the two related beliefs can generalize across different settings. Thus, some argue that it is the most robust, parsimonious, and influential model in explaining information technology adoption behaviour (Elliot & Loebbecke, 2000; Teo & Pok, 2003; Venkatesh, Morris, Davis, & Davis, 2003). Indeed, since its development, it has received extensive empirical support through validations, applications, and replications for its prediction power (Taylor & Todd, 1995, 1995a; Venkatesh & Morris, 2000a). A number of modified TAM models were proposed to suit new technologies including Internet and intranet (Agarwal & Prasad, 1998; Chau, 1996; Chau & Hu, 2001; Horton, Buck, Waterson, & Clegg, 2001). For example, TAM has been used to predict Internet purchasing behaviour (Gefen, Karahanna, & Straub, 2003; Kaufaris, 2002).

A major theoretical limitation of TAM is the "exclusion of the possibility of influence from institutional, social, and personal control factors" (Elliot & Loebbecke, 2000, p. 49). Thus the suitability of the model for predicting general individual acceptance needs to be re-assessed as the main TAM constructs do not fully reflect the specific influences of technological and usage-context factors that may alter user acceptance (King, Gurbaxani, Kraemer, McFarlan, Raman, & Yap, 1994; Taylor & Todd, 1995). In response to this, a number of modifications and changes to the original TAM models have been made. The most prominent of these is the unified theory of acceptance and use of technology (UTAUT), a unified model that integrates constructs across eight models (Venkatesh et al., 2003). UTAUT provides a refined view of how the determinants of intention and behaviour evolve over time and assumes that there are three direct determinants of intention to use (performance expectancy, effort expectancy, and social influence) and two direct determinants of usage behaviour (intention and facilitating conditions). However, both TAM and UTAUT have received criticisms with

the fundamental one being about the problems in applying these beyond the workplace and/or organisation for which originally created (Carlsson et al., 2005).

Motivational Theories

Motivation theories are rooted in psychological research to understand individuals' acceptance of information technology (Davis, Bagozzi, & Warshaw, 1992; Igbaria, Parasuraman, & Baroudi, 1996). These theories often distinguished extrinsic and intrinsic motivation. While extrinsic motivation refers to the performance of an activity in helping achieve valued outcomes, intrinsic motivation puts emphasis on the process of performing an activity (Calder & Staw, 1975; Deci & Ryan, 1985). For example, perceived usefulness is an extrinsic source of motivation (Davis et al., 1992) while perceived enjoyment (Davis et al., 1992), perceived fun (Igbaria et al., 1996), and perceived playfulness (Moon & Kim, 2001) can be considered intrinsic sources of motivation. Both sources of motivation affect usage intention and actual usage. Therefore, in addition to ease of use and usefulness, intrinsic motivators, such as playfulness, will also play an important role in increasing usability in a usage environment in which information technology applications are both used for work and play (Moon & Kim, 2001).

Innovation Diffusion Theory

The innovation diffusion theory is concerned with how innovations spread and consists of two closely related processes: the diffusion process and adoption process (Rogers, 1995). Diffusion is a macro process concerned with the spread of an innovation from its source to the public whereas the adoption process is a micro process that is focused on the stages individuals go through when deciding to accept or reject an innovation. Key elements in the entire process are the innovation's perceived characteristics, the individual's attitude and beliefs, and the communication received by individuals from their social environment. In relation to the factors pertaining to innovation, factors such as, relative advantage, complexity, trialability, observability, and compatibility, were considered important in influencing individual's acceptance of the innovation (Rogers, 1995).

TOWARDS AN ACCEPTANCE MODEL FOR MOBILE SERVICES

This section develops an acceptance model for mobile technology and services that may be empirically tested. This development begins with identifying the latent constructs in extant technology adoption literature. However, mobile services differ from traditional systems in that mobile services are ubiquitous, portable, and can be used to receive and disseminate personalised and localised information (Siau, Lim, & Shen, 2001; Teo & Pok, 2003). Thus, the models examined in the previous section and the constructs included in these models may not be applicable to mobile services adoption. In particular, we discuss the various antecedents of attitude towards mobile services and develop a new model based on the widely used TAM model to predict adoption of new mobile services.

User Predisposition

User predisposition refers to the internal factors of an individual user of mobile services. Personal differences strongly influence adoption. There is evidence that successful acceptance of innovations depends as much on individual adopter differences as on the innovation itself. Indeed, individual differences help identify segments of adopters who are more likely to adopt technology innovations than others, which in turn, helps providers address adopter needs more closely (Massey et al., 2005). Diffusion resources can also be used

more effectively and efficiently (Agarwal & Prasad, 1998). Early adopters, for example, can then act as opinion leaders or change agents to facilitate the diffusion of the technology further (Rogers, 1995). There are several dimensions used to capture individual differences, including personal innovativeness, perceived costs, demographic factors, psychographic profiles, and personality traits (Dabholkar & Bagozzi, 2002). In this chapter, we define user predisposition as the collection of factors including the individual's prior knowledge and experience of existing mobile services, compatibility, behavioural control, personal innovativeness, perceived enjoyment, and price sensitivity.

First, *prior knowledge* is essential for the comprehension of the technology and related services. According to Rogers (1995), knowledge occurs when a potential adopter learns about the existence of an innovation and gains some understanding concerning its functionality. Like other technologies, the mobile technology is comprised of both the hardware (i.e., the physical mobile device) and software domains (i.e., the applications consisting of the instructions to use the hardware as well as other information aspects) (Rogers, 1995). Thus knowledge from both hardware and software domains might be required for complete comprehension (Moreau, Lehmann, & Markman, 2001; Saaksjarvi, 2003). Prior knowledge consists of two components, namely, familiarity and expertise. For instance, the former constitutes the number of mobile services-related experiences accumulated by consumers over time, which includes exposure to advertising, information search, interaction with salespersons, and so on. The latter represents the ability to use the mobile services, and it includes beliefs about service attributes (i.e., cognitive structures) as well as decision rules for acting on those beliefs (i.e., cognitive processes) (Alba & Hutchinson, 1987). In any case, familiarity alone cannot capture the complexity of consumer knowledge (Alba & Hutchinson, 1987), which suggests the learning is required (Saaksjarvi, 2003).

With learning, consumers use the "familiar" component of existing knowledge as a means to understand and comprehend new phenomena in the innovation which is being adopted (Roehm & Sternthal, 2001). Specifically, existing knowledge in general and analogical learning in particular have been shown to be powerful and highly persuasive communication devices in acquiring in-depth understanding of innovation benefits and functionality (Moreau et al., 2001; Roehm & Sternthal, 2001; Yamauchi & Markman, 2000). An analogy compares and contrasts a known base innovation to an unknown target innovation. The base and the target share structural attributes, but are different in terms of surface attributes. A cellular phone versus a personal digital assistant (PDA) versus a "smart phone" are good examples. Research shows that "a message containing an analogy is better comprehended and is more persuasive when the recipient has expertise with regard to the base product [innovation]." (Roehm & Sternthal, 2001, p. 269). However, expertise alone is insufficient to ensure analogy persuasiveness. Substantial resources, training/usage instructions, and a positive mood are also required to facilitate learning (Roehm & Sternthal, 2001). However, while knowledge is important, by itself, it has limited usefulness, and therefore, "knowledge alone cannot determine the basis for adoption" (Rogers, 1995, p. 167) of a technology or service.

Adopters' previous positive or negative *experiences* with a technology or service can have a significant impact on their perceptions and attitudes towards that technology (Lee et al., 2003; Taylor & Todd, 1995a). Specifically, experience may influence adopters in forming positive or negative evaluations concerning innovations, which can boost or impair adoption of mobile technologies and services. Because of their greater clarity and certainty, direct prior experiences are likely to have a stronger impact on perceptions and attitudes towards usage than indirect or incomplete evidence (i.e., pre-trial) (Knutsen et al., 2005; Lee et al., 2003).

The second variable within the user predisposition construct is *compatibility*. Rogers (1995) defines compatibility as the degree to which an innovation is perceived to be consistent with existing values of potential adopters. In general, high incompatibility will adversely affect potential adopters of an innovation, which decreases the likelihood of adoption (Saaksjarvi, 2003). In contrast, high compatibility is likely to increase adoption propensity. In the context of wireless devices, lifestyle compatibility is the extent to which adopters believe mobile devices and services can be integrated into their daily lives. For example, adopters' lifestyle in terms of degree of mobility is likely to have a strong impact on their decision to adopt the technology (Pagani, 2004; Teo & Pok, 2003). For example, a person who leads a busy lifestyle, and is employed in an information-intensive job, and is always on the move is more likely to adopt a wireless device and its associated services compared to a person who leads a sedentary lifestyle.

Third, perceived *behavioural control*, a dynamic and socio-cognitive concept, has attracted a lot of attention in adoption literature. Earlier work by Ajzen (1991) considered it as a uni-dimensional variable. More recent empirical findings suggest that perceived behavioural control has two distinct components: self-efficacy, which is an individual's judgement of their capability to perform a behaviour, and controllability, which constitutes an individual's beliefs if they have the necessary resources and opportunities to adopt the innovation. It denotes a subjective judgment of the degree of control over the performance of a behaviour not the perceived likelihood that performing the behaviour will produce a given outcome (Ajzen, 1991). In the context of mobile service adoption, perceived behavioural control refers to the individual perception of how easy or difficult it is to get mobile services.

Fourth, *personal innovativeness* is the willingness of an individual to try out and embrace new technologies and their related services for accomplishing specific goals. Also known as technology readiness, personal innovativeness embodies the risk-taking propensity which exists in certain individuals and not in others (Agarwal & Prasad, 1998; Massey et al., 2005; Parasuraman, 2000). This definition helps segment potential adopters into what Rogers (1995) characterises as innovators, early adopters, early and late majority adopters, and laggards. Personal innovativeness represents a confluence of technology-related beliefs which jointly determine an individual's predisposition to adopt mobile devices and related services. The adoption of any innovation in general, and of innovative mobile phones and services in particular is inherently associated with greater risk (Kirton, 1976). Therefore, given the same level of beliefs and perceptions about an innovation, individuals with higher personal innovativeness are more likely to develop positive attitudes towards adopting it than less innovative individuals (Agarwal & Prasad, 1998).

Fifth, *perceived enjoyment* refers to the degree to which using an innovation is perceived to be enjoyable in its own right and is considered to be an intrinsic source of motivation (Al-Gahtani & King, 1999). Because the market for mobile innovations and services is comprised of both corporate users and consumers, factors focusing on perceived enjoyment constitute an important consideration (Carlsson et al., 2005; Pagani, 2004). That is, adopters use an innovation for the pleasure or enjoyment its adoption might bring and, therefore, serve as an end unto itself. Further, intrinsic enjoyment operates outside valued outcomes or immediate material needs (i.e., extrinsic motivations), such as enhanced job performance, increased pay, and so forth (Mathwick et al., 2001; Moon & Kim, 2001). Most research on enjoyment is based on the "flow theory" according to which flow represents "the holistic sensation that people feel when they act with total involvement" (Csikszentmihalyi, 1975). In a "flow state" individuals interact more voluntarily with innovations within their specific context, which determines their subjective expe-

riences (Csikszentmihalyi, 1975). Consequently, individuals who have a more positive enjoyment experience with an innovation are likely to have stronger adoption intentions than those who do not (Moon & Kim, 2001).

That is, intrinsic enjoyment can positively affect the adoption and use of innovative mobile services, and is therefore, a significant determinant of intention and attitude towards adoption (Kaufaris, 2002; Novak, Hoffman, & Yung, 2000). Further, upon adoption, individuals are more likely to use the mobile services that offer enjoyment more extensively than those which do not. As a consequence, perceived enjoyment is also seen to have a significant effect beyond perceived usefulness (Davis et al., 1989a). However, the complexity of a mobile innovation or service has a negative effect on perceived enjoyment, suggesting that the potential impact of enjoyment may not be fully realised (Igbaria et al., 1996).

The final variable that needs to be added to the existing technology adoption models is *price sensitivity*. In the original technology acceptance models, the costs of adopting an innovation were not considered to be a relevant construct because the actual users did not have to pay for the technology. In an organisational setting, the cost would be incurred by the organisation. However, in the context of individual private adoption, cost becomes a relevant factor. There is evidence showing that perceived financial resources required to adopt mobile technologies and services constitute a significant determinant of behavioural intention (Kleijen et al., 2004; Lin & Wang, 2005). However, evidence also shows that adopters of mobile devices and services also attempt to assess the value of adoption by comparing perceived costs against the benefits (Pagani, 2004). Perceived costs are directly related to income and socioeconomic status of potential adopters which are recognised to have a strong impact on technology adoption and diffusion (Lu, Yu, Liu, & Yao, 2003). For example, in Europe individuals earning income beyond certain levels were found to have a high propensity to embrace mobile technologies, such as WAP mobile phones, handheld computers, and so forth (Crawford, 2002). Similarly, there's evidence that in fast growing economies, individuals with higher income spend more on mobile devices (Lu et al., 2003).

Perceived Usefulness

Perceived usefulness is "the degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, 1989, p. 320). Perceived usefulness is also known as performance expectancy (Venkatesh et al., 2003). An innovation is believed to be of high usefulness when a potential adopter believes that there is a direct relationship between use on the one hand and productivity, performance, effectiveness, or satisfaction on the other (Lu et al., 2003).

Usefulness recognition is important because it has been found to have a strong direct effect on the intention of adopters to use the innovation (Adams, Nelson, & Todd, 1992; Davis, 1989). In addition, potential adopters assess the consequences of their adoption behaviour and innovation usage in terms of the ongoing desirability of usefulness (Chau, 1996; Venkatesh & Davis, 2000). Although an innovation might provide at least some degree of usefulness, a potential reason not to adopt exists when adopters fail to see the "need" to adopt (Zeithaml & Gilly, 1987). Adopters may not be able to recognise their needs until they become aware of the innovation or its consequences (Rogers, 1995). Need recognition is, therefore, likely to drive potential adopters to educate themselves in order to be able to utilise the innovation fully before being able to recognise its usefulness. This in turn is likely to lead to a faster rate of adoption (Rogers, 1995; Saaksjarvi, 2003).

Perceived usefulness can be split into two parts. Near-term usefulness is perceived to have an impact on the near-term job fit, such as job performance or satisfaction (Thompson, Higgins, & Howell, 1994). Long-term usefulness is

perceived to enhance the future consequences of adoption including career prospects, opportunity for preferred job assignments, or social status of adopters (Chau, 1996; Thompson et al., 1994). Evidence shows that even though perceived near-term usefulness has the most significant impact on the behavioural intention to adopt an innovation, perceived long-term usefulness also exerts a positive, yet lesser impact (Chau, 1996; Jiang, Hsu, Klein, & Lin, 2000).

In the case of mobile technology and services, perceived usefulness is defined as the degree to which the mobile technology and services provide benefits to individuals in every day situations (Knutsen et al., 2005). The range and type of service offerings as well as the compatibility of the user's existing computing devices influence perceived usefulness (Pagani, 2004). In addition, Pagani (2004) also finds that usefulness emerges as the strongest determinant in the adoption of three generation mobile services which is consistent finds of research concerning the adoption of other innovations (Venkatesh et al., 2003).

Perceived Ease of Use

Perceived ease of use is the "degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320). Other constructs that capture the notion of perceived ease of use are complexity and effort expectancy (Rogers, 1995; Venkatesh et al., 2003). Perceived ease of use may contribute towards performance, and therefore, near-term perceived usefulness. In addition, lack of it can cause frustration, and therefore, impair adoption of innovations. Nevertheless, "no amount of EOU [ease of use] will compensate for low usefulness" (Keil, Beranek, & Konsynski, 1995, p. 89).

In the mobile setting, perceived ease of use represents the degree to which individuals associate freedom of difficulty with the use of mobile technology and services in everyday usage (Knutsen et al., 2005). For example, there is evidence in the media that using certain services on a mobile device can be quite tedious, especially when browsing Internet-like interfaces on mobile devices is required (Teo & Pok, 2003). Together with relatively small screen sizes and associated miniaturized keypads, the overall usage experience may be adversely affected. This suggests that input and output devices are likely to influence perceived ease of use (Pagani, 2004). In addition, user-friendly and usable intuitive man-machine interfaces, including clear and visible steps, suitable content and graphical layouts, help functions, clear commands, symbols, and meaningful error messages are likely to influence adoption as well (Condos, James, Every, & Simpson, 2002). Further, Pagani (2004) argues the mobile system response time affects perceived ease of use suggesting that mobile bandwidth is important as well.

Social Influences

Social influence constitutes the degree to which individuals perceive that important or significant others believe they should use an innovation (Venkatesh et al., 2003). Venkatesh et al. (2003) believe that the social influence constructs may only become significant drivers on intention to adopt when users adopt an innovation in order to comply mandatory requirements. In these circumstances, social influence seems to be significant in the early phases of adoption and its effect decreases with sustained usage (Venkatesh & Davis, 2000). Conversely, in voluntary settings, social influence appears to have an impact on perceptions about the innovation (Venkatesh et al., 2003). Social influence is related to three similar constructs, namely, subjective and social norms, and image.

In Taylor and Todd's study (1995), subjective norms are defined to include the influence of other people's opinions otherwise known as reference groups. These include peers, friends, superiors, computer, and technology experts. Subjective

norms have a greater impact during the initial adoption phase when potential adopters have little or no experience or when the adoption behaviour is new (Thompson et al., 1994). Research shows that pressure from reference groups to adopt an innovation is effective because it contributes to reducing perceived risk associated with adoption (Teo & Pok, 2003).

Social factors constitute another construct of social influence. Social factors represent cues individuals receive from members of their social structure which prompt them to behave in certain ways (Thompson, Higgins, & Howell, 1991). For example, in Japan, teenagers regard smart phones as fashion items (Lu et al., 2003). Further, there is evidence that unique communications patterns determined by key social and cultural factors, such as group-oriented nationality, have positively affected adoption practices of using the Internet via mobile phones in East Asia (Ishii, 2004).

A third critical construct related to social influence is image. The adoption of an innovation can be seen to enhance one's status or image in their social system. For certain adopters a mobile device may be more of a lifestyle than a necessity (Bina & Giaglis, 2005; Teo & Pok, 2003). For example, early adopters of mobile computing devices might be image-conscious users who wish to be seen as trend-setters or technology savvy enthusiasts.

Facilitating Conditions

Facilitating conditions refer to external controls and catalysts in the adoption environment which aim at facilitating adoption and diffusion of new technologies (Terry, 1993). Facilitating conditions are important because they are considered to be direct usage antecedents, and are therefore, likely to make adoption behaviour less difficult by removing any obstacles to adoption and sustained usage (Thompson et al., 1994; Venkatesh et al., 2003). These conditions can be provided by both governments and mobile operators. For example, governments or the representative agencies can act as facilitators by bringing together the telecommunication industry, academia, and research community. Government agencies can also set up protocol standardization policies and regulations favouring the future growth of mobile communication systems (Lu et al., 2003). Likewise, mobile operators can encourage adoption by mass advertising campaigns and active promotion aimed at increasing awareness about mobile devices and related services (Teo & Pok, 2003). Further, promoting and enforcing appropriate interconnection agreements and adequate regulatory mechanisms among mobile operators help adopters of mobile devices take advantage of roaming services and consequently be conducive to adoption (Rossotto et al., 2000).

Facilitating conditions also capture the existence of a trusting environment that is external to the mobile operator's control. A trusting environment constitutes an important factor in the adoption of mobile technologies and services. It determines the user's expectations from the relationship with their service providers, and it increases their perceived certainty concerning the provider's expected behaviour. Generally, trust is essential in all economic activities where undesirable opportunistic behaviour is likely to occur (Gefen et al., 2003). However, trust becomes vital in a mobile environment, where situational factors such as uncertainty or risk and information asymmetry are present (Ba & Pavlou, 2002). On the one hand, adopters of mobile technology are unable to judge the trustworthiness of service providers, and on the other, the latter can also easily take advantage of the former by engaging in harmful opportunistic behaviours. For example, service providers can sell or share the transactional information of its users or their personal information.

There are two key elements in a trusting environment, namely, security and privacy (Lu et al., 2003). In a wireless environment, security encompasses confidentiality, authentication, and message integrity. Because mobile devices

have limited computing resources and wireless transmissions are more susceptible to hacker attacks, security vulnerabilities can have serious consequences (Galanxhi-Janaqi & Nah, 2004; Lu et al., 2003). There are several remedies against the dangers of insecurity, for example, public key infrastructure and certificate authority which use public key cryptography to encrypt and decrypt mobile transmissions and authenticate users.

Ironically, the same information practices which provide value to both users and providers of mobile technology and services also cause privacy concerns. Some of these concerns include: the type of information that can be collected about users and the ways in which it will be protected; the entities that can access this information and their accountability; and the ways in which the information will be used (Galanxhi-Janaqi & Nah, 2004). In mobile adoption research the trust environment has been encapsulated in a construct called perceived credibility (Lin & Wang, 2005; Wang, Wang, Lin, & Tang, 2003). Evidence shows that there is a "significant direct relationship between perceived credibility and behavioural intention" (Lin & Wang, 2005, p. 410) to use mobile services.

Moderating Variables

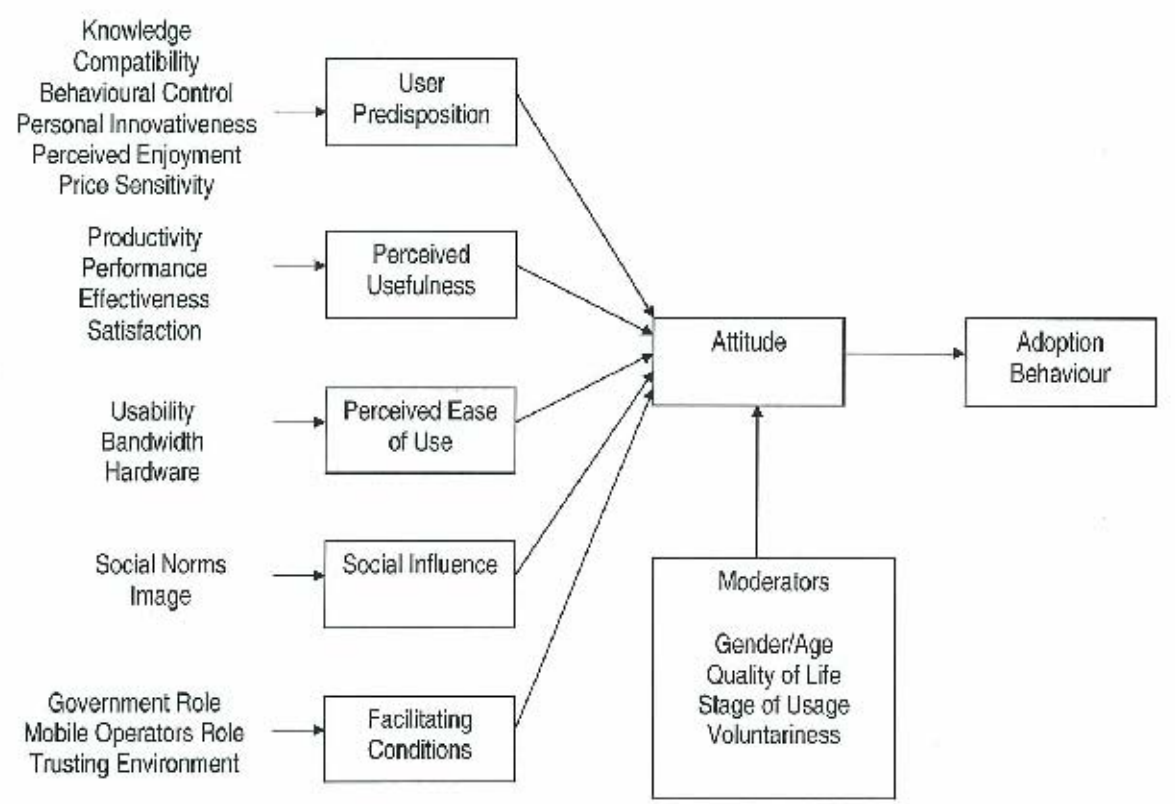
Evidence shows that gender and age might influence the adoption of technology and related services due to their moderating effects on other constructs (Venkatesh et al., 2003). In general, men tend to exhibit task-oriented attitudes suggesting that usefulness expectations might be more accentuated in men than women (Minton & Scheneider, 1980). This is particularly the case for younger men (Venkatesh & Morris, 2000a). On the other hand, ease of use expectations are more salient for women and older adopters (Bozionelos, 1996). Further, women are predisposed to be more sensitive to the opinions of members of their social structure. As a result women are more likely to be affected by social influence factors when deciding to adopt new mobile technologies and services (Venkatesh & Morris, 2000a). Similarly, because affiliation needs increase with age (Rhodes, 1983), older adopters are more likely to be affected by social influence.

Quality of life of potential adopters is another moderating variable which is likely to affect the adoption of mobile devices and services. "Quality of life" is an established social sciences notion which represents "a global assessment of a person's life satisfaction according to his chosen criteria" (Diener & Suh, 1997; Shin & Johnson, 1978). There is evidence which indicates that mobile technology and services have enhanced the perceived quality of social and work life of adopters (Jarvenpaa, Lang, Takeda, & Tuunainen, 2003). Bina and Giaglis (2005) present evidence that the reverse is also true. They indicate that adopters who are satisfied with specific life domains exhibit favourable attitudes towards the adoption of specific mobile services (Bina & Giaglis, 2005).

Further, evidence also shows that stage of use and voluntariness of usage have moderating effects on adoption attitudes through various constructs. For example, perceived ease of use is significant during the initial period of usage when process-related issues constitute obstacles to be overcome (Venkatesh et al., 2003). Perceived ease of use, however, becomes insignificant during periods of extended usage (Agarwal & Prasad, 1998; Davis et al., 1989a; Thompson et al., 1994). Similarly, ease of use has a significant positive effect on attitude toward use in both voluntary and mandatory usage contexts (Al-Gahtani & King, 1999; Keil et al., 1995; Venkatesh & Davis, 2000).

To summarise the constructs discussed in this section, Figure 2 portrays a proposed model of acceptance of mobile services. The implications of this model are discussed next.

Figure 2. Proposed model of acceptance of mobile services



PROPOSED METHOD

In order to validate the model discussed in the previous section, we propose a two-stage research design, consisting of both qualitative and quantitative approaches.

Qualitative Stage

This stage is the first round of the fieldwork data collection. Data collection at this stage would involve conducting face-to-face in-depth interviews in order to study the perceptions of all stakeholders who contribute directly or indirectly to providing mobile services. Stakeholder targets include mobile operators or carriers, industry and government associations, user groups, mobile application developers, content providers, aggregators, as well as manufacturers of mobile devices. The aim of these interviews is to gather an in-depth understanding of the perceptions and perspectives of all stakeholders involved in the adoption and diffusion of mobile services.

To ensure validity in this research design, several tactics may be used. Construct validity would be addressed by using the multiple sources of evidence as noted previously. The issue of internal validity would be considered by using the techniques of pattern matching the data to a predicted pattern of variables, and formulating rival explanations. In addition, an interview protocol should be developed to guide the data collection. A single pilot case study is recommended to be used in order to refine data collection procedures and improve conceptualisation of the model prior to finalising

the set of theoretical propositions developed from the literature.

The second round of the qualitative stage would involve focus groups with mobile users. Because the research phenomenon is contemporary and little prior research has been conducted, focus groups would be appropriate for generating ideas and obtaining insights from existing mobile service users and potential users (Carson, Gilmore, Gronhaug, & Perry, 2001). Focus groups are useful when investigating complex behaviour and motivations. By comparing the different points of view that participants exchange during the interactions in focus groups, researchers can examine motivation with a degree of complexity that is usually not available with other methods (Morgan & Krueger, 1993). The use of a focus group is more valuable many times over compared with any representative sample for situations requiring the investigation of complex decision-making processes, as is the case for this research. Based on demographic characteristics, we propose setting up homogeneous groups because discussions within homogeneous groups produce more in-depth information than discussions within heterogenous groups (Bellenger, Bernhardt, & Goldtucker, 1989). These groups would be selected based on main moderating variables identified in the literature, such as, gender/age, quality of life, stage of usage, and voluntariness.

We believe that at least two investigators should conduct all interviews and moderate the focus groups (Denzin, 1989; Patton, 1990). This kind of triangulation reduces the potential bias which is commonly cited as a limitation of interviews and focus gruops (Frankfort-Nachmias & Nachmias, 1996; Yin, 1994).

Quantitative Stage

The last stage of this project would involve an online survey. The collected data would help understand and confirm the determinants and the adoption intentions of the consumers of mobile services. Random sampling should be used to select the sample. We propose that two types of data analysis should be performed on the survey data: descriptive analysis and inferential analysis. Descriptive analysis should be carried out for transformation of raw data into a form that would provide information to describe a set of factors in a situation (Sekaran, 2000). For the inferential analysis, a structural equation model (SEM) should be used to test the refined model.

MANAGERIAL IMPLICATIONS

Mobile technologies and the associated services integrate both the business and social domains of the user's life (Elliot & Phillips, 2004; Knutsen et al., 2005). 3G services in general and location-based services in particular can provide anytime-anyplace tracking of adopters (UMTS, 2003). This creates the opportunity for developing accurate adopter profiles both in their work- and leisure-related domains. In addition, live video and location-based information can also be gathered (Robins, 2003). While such information can help address the needs of adopters better, it can also be misused by businesses for unethical direct business-to-consumer marketing (Casal et al., 2004), raising privacy concerns, overcontrol and overwork of individual adopters (Yen & Chou, 2000). For example, by reducing space and time constraints, mobile communications provide an immensely flexible work environment for some individuals while bringing about overwork or intrusion problems for others (Gerstheimer & Lupp, 2004). As a result, existing privacy protection policies and regulations about employees and consumers should reflect these new conditions. These policies should also account for overcontrol prevention that is likely to result from organisations' attempts to monitor individual performance (Yen & Chou, 2000).

Designing content suitable for mobile phones constitutes an important issue that affects the adoption and diffusion of mobile technology and associated services. This has implications

for service providers, developers, policymakers, and academics. Content providers must design content "for value-contexts specific for mobile use which provide users freedom from complicated configuration procedures, and ubiquitously serve and support current day-to-day individual social practices" (Knutsen et al., 2005a, p. 7). Developers of mobile applications need to recognise that mobile applications are quite different from PC applications (Funk, 2005). Developers should use established standards, such as HTML and Java. More importantly, the usage contexts, the adopters, and their evolving behaviour should be important considerations. Further, because "made-for-the-medium" content type and design may be required (Massey et al., 2005), the available technologies which determine screen size, display quality and processing speeds should be taken into consideration as well (Funk, 2005). The combined effect of these factors on navigation patterns, adopters' cognitive overload, and subjective perceptions about the usability and ease of use of mobile applications can have a critical impact on uptake (Chae & Kim, 2004).

Segmentation of mobile service adopters must not only be based on adopter type (e.g., pioneers, early adopters, majority adopters, and laggards) but also on individual differences. The basis of segmentation should constitute the foundation in developing marketing strategies. For example, individuals with high personal innovativeness or novelty seekers are likely to be willing to experiment with new mobile devices and services, in which case these should be marketed as technological innovations. For individuals who are reluctant to use the same devices and services and are likely to feel discomfort and insecurity while using them, lifestyle promotions may be more appropriate (Dabholkar & Bagozzi, 2002; Teo & Pok, 2003). In addition, endorsements by peers, famous celebrities, or other referent groups may be adequate if these individuals appreciate social norms and image (Hung, Ku, & Chang, 2003; Teo & Pok, 2003). Marketing mobile applications for adopters in one category is likely to frustrate adopters in the other. Therefore, developers and marketers should be prudent in recognising that the confluence of various individual characteristics with varying levels of prior experience, perceptions, and learning predispositions are all likely to influence adoption and retention patterns (Card, Moran, & Newell, 1983; Hung et al., 2003; Massey et al., 2005).

Further, the interface design of mobile applications should encompass both intrinsic and extrinsic motivation dimensions (Moon & Kim, 2001). Based on the proposed model, marketers should promote attributes such as usefulness, ease of use, and enjoyment as important aspects when attempting to persuade potential users in adopting specific mobile phones and services as well as to increase their loyalty and retention (Dabholkar & Bagozzi, 2002; Hung et al., 2003; Lin & Wang, 2005). In particular, personalisation is a well-suited and an achievable goal as mobile phones are identifiable. 3G phones also enable identification of the location of individual handsets, making location-specific marketing possible. Messages promoting the services of businesses, such as restaurants, hotels, grocery stores, and so forth, can be transmitted when users are detected within range (Robins, 2003). Evidence shows that despite privacy concerns, many users of mobile devices are happy to receive unsolicited promotional messages provided that such messages are relevant and personalised (Robins, 2003).

Governments and mobile operators should design appropriate and dedicated strategies to promote the relative advantages of mobile phones and services. Such promotion strategies are important because of their impact on the perceptions of potential adopters (Knutsen et al., 2005). Moreover, the development of wireless communication infrastructures and the provision of incentives are likely to contribute towards the minimisation of the digital divide which results from demographic factors such as varying income levels, education and experience, gender, and age (Lin & Wang, 2005). The digital divide not only prevents the exploitation of the full market potential, but it also

adversely impacts the maximization of benefits for current adopters due to limited network externalities effects (Katz & Shapiro, 1986).

CONCLUSIONS AND FUTURE RESEARCH

User acceptance of mobile technology and related services is of paramount importance. Consequently, a deeper insight into theory-based research is required to better understand the underlying motivations and barriers that will lead users to inhibit them from adopting these technologies and services. This in turn will also help designing technology and service improvements as well as appropriate adoption and diffusion strategies. There are several theoretical models in the literature which attempt to determine acceptance and adoption of new technologies. However, most of these models originate from organisational contexts. As mobile technologies and services add other functional dimensions such as hedonic or experiential aspects, applying extant theories outright to determine the acceptance and adoption of mobile services may be questionable and inadequate.

In this chapter, we have explored and critically reviewed existing technology acceptance theories. Relevant constructs of extant models were discussed in the light of evolving mobile technologies and services and then incorporated into a synthesised acceptance model of mobile services. The proposed model attempts to view acceptance of mobile services beyond traditional organisational borders and permeate everyday social life practices. The proposed model which can be tested empirically provides the foundation to guide further validation and future research in the area of mobile services adoption.

In addition, a plethora of mobile services have become available recently (Alahuhta et al., 2005). Because all services would be available to adopters through a single user interface of the current technology, the appropriation of these services by users may be interconnected and at different stages of maturity (Knutsen et al., 2005). These interconnections are temporal and are also likely to have mutually enhancing, suppressing, or compensating effects on each other (Black & Boal, 1994). This adds dynamism and complexity to acceptance and, therefore, cannot be explained by simply considering factors impacting individual or aggregate adoption at single points in time (Knutsen et al., 2005; Pagani, 2004). Consequently, future research should develop and test dynamism-compatible acceptance models because these models may provide a deeper understanding and help in explaining how and why technology acceptance perceptions change as the appropriation process progresses. Further, with a wide variety of mobile devices and services available and their applicability in distinct spheres of life, the definition of a unit of analysis in mobile services adoption has become a challenging task (Knutsen et al., 2005). Additional research in this aspect is also needed.

REFERENCES

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly, 16*(2), 227-247.

Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research, 9*(2), 204-215.

Ajzen, I. (1991). The theory of planned behavior. *Organisational Behavior and Human Decision Process, 52*(2), 179-211.

Alahuhta, P., Ahola, J., & Hakala, H. (2005). *Mobilising business applications: a survey about the opportunities and challenges of mobile business applications and services in Finland* (Technology Review No. 167/2005). Helsinki: Tekes.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research, 13*(3), 411-454.

Al-Gahtani, S. S., & King, M. (1999). Attitudes, satisfaction and usage: Factors contributing to each in the acceptance of information technology. *Behaviour & Information Technology, 18*(4), 277-297.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. *MIS Quarterly, 26*(3), 243-268.

Bellenger, D. N., Bernhardt, K. L., & Goldtucker, J. L. (1989). Qualitative research techniques: Focus group interviews. In T. J. Hayes, & C. B. Tathum (Eds.), *Focus group interviews: A reader* (pp. 7-28). Chicago: American Marketing Association.

Bina, M., & Giaglis, G. M. (2005). *Exploring early usage patterns of mobile data services.* Paper presented at the International Conference on Mobile Business, Sydney, Australia, July 11-13.

Black, J. A., & Boal, K. B. (1994). Strategic resources: Traits, configurations, and paths to sustainable competitive advantage. *Strategic Management Journal, 15*, 131-148.

Bozionelos, N. (1996). Psychology of computer use: Prevalence of computer anxiety in British managers and professionals. *Psychological Reports, 78*(3), 995-1002.

Calder, B. J., & Staw, B. M. (1975). Self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology, 31*(4), 599-605.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Carlsson, C., Hyvonen, K., Repo, P., & Walden, P. (2005). *Adoption of mobile services across different platforms.* Paper presented at the 18th Bled eCommerce Conference, Bled, Slovenia, June 6-8.

Carson, D., Gilmore, A., Gronhaug, K., & Perry, C. (2001). *Qualitative research in marketing.* London: Sage.

Casal, C. R., Burgelman, J. C., & Bohlin, E. (2004). Propects beyond 3G. *Info, 6*(6), 359-362.

Chae, M., & Kim, J. (2003). What's so different about the mobile Internet? *Communications of the ACM, 46*(12), 240-247.

Chae, M., & Kim, J. (2004). Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour & Information Technology, 23*(3), 165-181.

Chau, P. Y. K. (1996). An empirical assessment of a modified technology acceptance model. *Journal of Management Information Systems, 13*(2), 185-204.

Chau, P. Y. K., & Hu, P. J.-H. (2001). Information technology acceptance by individual professionals: a model comparison approach. *Decision Science, 32*(4), 699-719.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly, 23*(2), 189-211.

Condos, C., James, A., Every, P., & Simpson, T. (2002). Ten usability principles for the development of effective WAP and m-commerce services. *Aslib Proceedings, 54*(6), 345-355.

Crawford, A. M. (2002). International media habits on the rise. *Ad Age Global, 2*(11). Retrieved from http://web.ebscohost.com/ehost/detail?vid=3&hid=101&sid=ff86c2ae-e7f7-4388-96b4-7da9c1bc4eb3%40sessionmgr106

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety.* San Francisco: Jossey-Bass.

Dabholkar, P. A., & Bagozzi, R. P. (2002). An attitudinal model of technology-based self-service: Moderating effects of consumer traits and situational factors. *Journal of Academy of Marketing Science, 30*(3), 184-201.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance in information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1002.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology, 22*, 1111-1132.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum Press.

Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods (3rd ed.).* Englewood Cliffs, N. J.: Prentice Hall.

Dholakia, R. R., & Dholakia, N. (2004). Mobility and markets: Emerging outlines for m-commerce. *Journal of Business Research, 57*(12), 1391-1396.

Diener, E., & Suh, E. (1997). Measuring quality of life: Economic, social and subjective indicators. *Social Indicators Research, 40*(1-2), 189-216.

Dillon, A., & Morris, M. (1996). User acceptance of information technology: theories and models. *Journal of American Society for Information Science, 31*, 3-32.

Elliot, G., & Phillips, N. (2004). *Mobile commerce and wireless computing systems.* Harlow: Pearson Education Limited.

Elliot, S., & Loebbecke, C. (2000). Interactive, inter-organizational innovations in electronic commerce. *Information Technology & People, 13*(1), 46-66.

Figge, S. (2004). Situation-dependent services: A challenge for mobile operators. *Journal of Business Research, 57*(12), 1416-1422.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences (5th ed.).* New York: St. Martin's Press.

Funk, J. L. (2005). The future of the mobile phone Internet: An analysis of technological trajectories and lead users in the Japanese market. *Technology in Society, 27*(1), 69-83.

Galanxhi-Janaqi, H., & Nah, F. F.-H. (2004). U-commerce: Emerging trends and research issues. *Industrial Management & Data Systems, 104*(9), 744-755.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: an integrated model. *MIS Quarterly, 27*(1), 51-90.

Gerstheimer, O., & Lupp, C. (2004). Needs versus technology: The challenge to design third-generation mobile applications. *Journal of Business Research, 57*(12), 1409-1415.

Grundström, C., & Wilkinson, I. F. (2004). The role of personal networks in the development of industry standards: A case study of 3G mobile telephony. *Journal of Business and Industrial Marketing, 19*(4), 283-293.

Hammond, K. (2001). B2C e-commerce 2000-2010: What experts predict. *Business Strategy Review, 12*(1), 43-50.

Hart, J., & Hannan, M. (2004). The future of mobile technology and mobile wireless computing. *Campus-Wide Information Systems, 21*(5), 201-204.

Horton, R. P., Buck, T., Waterson, P. E., & Clegg, C. W. (2001). Explaining intranet use with the technology acceptance model. *Journal of Information Technology, 16*, 237-249.

Hung, S.-Y., Ku, C.-Y., & Chang, C.-M. (2003). Critical factors of WAP services adoption: An empirical study. *Electronic Commerce Research and Applications, 2*(1), 42-60.

Igbaria, M., Parasuraman, S., & Baroudi, J. J. (1996). A motivational model of microcomputer usage. *Journal of Management Information Systems, 13*(1), 127-143.

Ishii, K. (2004). Internet use via mobile phone in Japan. *Telecommunications Policy, 28*(1), 43-58.

Jarvenpaa, S. L., Lang, K. R., Takeda, Y., & Tuunainen, V. K. (2003). Mobile commerce at crossroads. *Communications of the ACM, 46*(12), 41-44.

Jiang, J. J., Hsu, M. K., Klein, G., & Lin, B. (2000). E-commerce user behaviour model: An empirical study. *Human Systems Management, 19*(4), 265-276.

Katz, M. L., & Shapiro, C. (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy, 94*(4), 822-841.

Kaufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behaviour. *Information Systems Research, 13*(2), 205-223.

Keil, M., Beranek, P. M., & Konsynski, B. R. (1995). Usefulness and ease of use: Field study evidence regarding task considerations. *Decision Support Systems, 13*(1), 75-91.

Khalifa, M., & Cheng, S. K. N. (2002). *Adoption of mobile commerce: Role of exposure*. Paper presented at the 35th Hawaii International Conference on System Sciences, Hilton Waikoloa Village, Hawaii, January 7-10 (pp. 46-52). IEEE Computer Society.

King, J. L., Gurbaxani, V., Kraemer, K. L., McFarlan, F. W., Raman, K. S., & Yap, C. S. (1994). Institutional factors in information technology innovation. *Information Systems Research, 5*(2), 139-169.

Kirton, M. (1976). Adopters and innovators: a description and measure. *Journal of Applied Psychology, 61*(5), 622-629.

Klasen, L. (2002). Migrating an online service to WAP: Case study. *The Electronic Library, 20*(3), 195-201.

Kleijen, M., Wetzels, M., & de Ruyter, K. (2004). Consumer acceptance of wireless finance. *Journal of Financial Services Marketing, 8*(3), 206-217.

Knutsen, L., Constantiou, I. D., & Damsgaard, J. (2005). *Acceptance and perceptions of advanced mobile services: Alterations during a field study.* Paper presented at the International Conference on Mobile Business, Sydney, Australia, July 11-13.

Lee, M. S. Y., McGoldrick, P. J., Keeling, K. A., & Doherty, J. (2003). Using ZMET to explore barriers to the adoption of 3G mobile banking services. *International Journal of Retail & Distribution Management, 31*(6), 340-348.

Leonard-Barton, D., & Deschamps, I. (1988). Managerial influence in the implementation of new technology. *Management Science, 34*(10), 1252-1265.

Lin, H., & Wang, Y. (2005). *Predicting consumer intention to use mobile commerce in Taiwan.* Paper presented at the International Conference on Mobile Business, Sydney, Australia, July 11-13.

Lu, J., Yu, C., Liu, C., & Yao, J. E. (2003). Technology acceptance model for wireless Internet. *Internet Research: Electronic Networking Applications and Policy, 13*(3), 206-222.

Massey, A. P., Khatri, V., & Ramesh, V. (2005). *From the Web to the wireless Web: Technology readiness and usability.* Paper presented at the 38th Hawaii International Conference on System Sciences, Hilton Waikoloa Village, Hawaii, January 3-6 (p. 32b). IEEE Computer Society.

Mathwick, C., Malhotra, N., & Rigdon, E. (2001). Experiental value: Conceptualization, measurement and application in the catalog and Internet shopping environment. *Journal of Retailing, 77*(1), 39-56.

Minton, G. C., & Scheneider, F. W. (1980). *Differential psychology.* Prospect Heights, IL: Waveland Press.

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management, 38*(4), 217-230.

Moreau, C. P., Lehmann, D. R., & Markman, A. B. (2001). Entrenched knowledge structures and consumer response to new products. *Journal of Marketing Research, 38*(1), 14-29.

Morgan, D. L., & Krueger, R. A. (1993). When to use focus groups and why. In D. L. Morgan (Ed.), *Successful focus groups* (pp. 1-19). London: Sage Publications.

Novak, T. P., Hoffman, D. L., & Yung, Y. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science, 19*(1), 22-42.

Pagani, M. (2004). Determinants of adoption of third generation mobile multimedia services. *Journal of Interactive Marketing, 18*(3), 46-59.

Parasuraman, A. (2000). Technology readiness index: A multiple item scale to measure readiness to embrace new technologies. *Journal of Service Research, 2*(4), 307-320.

Patton, M. Q. (1990). *Qualitative evaluation and research methods (2nd ed.).* London: Sage Publications.

Ratliff, J. M. (2002). NTT DoCoMo and its i-mode success: Origins and implications. *California Management Review, 44*(3), 55-71.

Repo, P., Hyvonen, K., Pantzar, M., & Timonen, P. (2004). *Users intenting ways to enjoy new mobile services: The case of watching mobile videos.* Paper presented at the 37th Hawaii International Conference on System Sciences, Hawaii, January 5-8 (p. 40096.3). IEEE Computer Society.

Rhodes, S. R. (1983). Age-related differences in work attitudes and behavior: A review of conceptual analysis. *Psychological Bulletin, 93*(2), 328-367.

Robins, F. (2003). The marketing of 3G. *Marketing Intelligence & Planning, 21*(6), 370-378.

Roehm, M. L., & Sternthal, B. (2001). The moderating effect of knowledge and resources on the persuasive impact of analogies. *Journal of Consumer Research, 28*(2), 257-272.

Rogers, E. M. (1995). *Diffusion of innovations.* New York: Free Press.

Rossotto, C. M., Kerf, M., & Rohlfs, J. (2000). Competition in mobile telecommunications: Sector growth, benefits for the incumbent and policy trends. *Info, 2*(1), 67-73.

Saaksjarvi, M. (2003). Consumer adoption of technological innovations. *European Journal of Innovation Management, 6*(2), 90-100.

Sekaran, U. (2000). *Research methods for business: A skill building approach*. New York: John Wiley and Sons.

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*(3), 325-343.

Shin, C. C., & Johnson, D. M. (1978). Avowed happiness as an overall assessment of quality of life. *Social Indicators Research, 5*, 475-492.

Siau, K., Lim, E. P., & Shen, Z. (2001). Mobile commerce: Promises, challenges, and research agenda. *Journal of Databases Management, 12*(2), 4-13.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research, 6*(2), 144-176.

Taylor, S., & Todd, P. A. (1995a). Assessing IT usage: The role of prior experience. *MIS Quarterly, 19*(4), 561-570.

Teo, T. S. H., & Pok, S. H. (2003). Adoption of WAP-enabled mobile phones among Internet users. *Omega: The International Journal of Management Science, 31*(6), 483-498.

Terry, D. J. (1993). Self-efficacy expectancies and the theory of reasoned action. In D. C. Terry, C. Gallois, & M. McCamish (Eds.), *The theory of reasoned action: Its application to AIDS-preventive behaviour* (pp. 135-152). Oxford: Pergamon.

Thompson, R., Higgins, C., & Howell, J. (1994). Influence of experience on personal computer utilization: Testing a conceptual model. *Journal of Management Information Systems, 11*(1), 167-187.

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly, 15*(1), 125-143.

UMTS. (2003). *Mobile evolution: Shaping the future*. Retrieved August 28, 2005, from http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Papers_index

van Steenderen, M. (2002). Business applications of WAP. *The Electronic Library, 20*(3), 215-223.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science, 46*(2), 186-204.

Venkatesh, V., & Morris, M. G. (2000a). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115-139.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478.

Wang, Y.-S., Wang, Y.-M., Lin, H.-H., & Tang, T.-I. (2003). Determinants of user acceptance of Internet banking: An empirical study. *International Journal of Service Industry Management, 14*(5), 501-519.

Xylomenos, G., & Polyzos, G. C. (2001). Quality and service support over multi-service wireless Internet links. *Computer Networks, 37*(5), 601-615.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 776-795.

Yen, D. C., & Chou, D. C. (2000). Wireless communications: Applications and managerial issues. *Industrial Management & Data Systems, 100*(9), 436-443.

Yin, R. K. (1994). *Case study research: Design and methods.* Beverley Hills: Sage.

Zeithaml, V. A., & Gilly, M. C. (1987). Characteristics affecting the acceptance of retailing technologies: A comparison of elderly and nonelderly consumers. *Journal of Retailing, 63*(1), 49-68.

This work was previously published in Mobile Multimedia Communications: Concepts, Applications, and Challenges, edited by G. Karmakar and L. Dooley, pp. 85-108, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.17
Multimedia Databases

Mariana Hentea
Southwestern Oklahoma State University, USA

INTRODUCTION

The Internet and technologies such as high-capacity storage devices, broadband telecommunications systems, and multimedia development software systems have accelerated the development of new applications based on the use of multimedia database systems. A multimedia database is a repository of different data objects (text, numeric values, Boolean values, dates, graphical images, video clips, and sound files). Examples of applications based on multimedia databases include news and stock market information on demand; movies on demand; home shopping; medical systems; trademark, patent, and copyright databases; geographic information systems (GIS); weather forecasting; Computer Aided Design (CAD) systems; architectural design; fabric and fashion design; interior design; photographic libraries; art gallery and museum management; law enforcement; criminal investigations; military reconnaissance and surveillance; scientific experiments; and educational systems.

The real-time nature and different kinds, including the size of multimedia data, cause problems for the design, implementation, and management of multimedia databases. Multimedia data requires use of specific technologies to reduce the size of the media data so it could be stored within the database. For example, NASA's Earth Observation System generates a terabyte (1,000 gigabytes) of data a day. This data comprises images recorded from orbiting satellites by video and infrared cameras that are downloaded to earth. The software chosen by the National Library of Australia is the TeraText™Database System, providing a single access point to over 600,000 images from 18 Australian cultural institutions, including libraries, museums, archives, and galleries representing images related to Australia's cultural heritage from the late 18th century through to the present day. The TeraText™Database System is known for its scalability, flexibility, speed, and sophistication. The increasing importance of multimedia applications and the introduction of SQL3 in 1999 have favored changes to traditional relational databases. These resulted in the implementation of features to store and manipulate large objects within object-relational databases. One example is Oracle9i, which is an object-re-

lational database management system that offers capabilities for holding media data. Through the use of media data types, Oracle interMedia software enables the Oracle9i database management system to manage and deliver image, audio, video, and geographical location data in an integrated fashion with other enterprise information.

The creation, storage, and data modeling techniques for multimedia data objects and retrieval of objects are main issues in the development of multimedia databases. Other issues include authenticity, integrity, media metadata, automatic video capture and editing, mobile media applications, scheduling, and quality of multimedia delivery over a network (Bourbakis, 2004; Shu & Yu, 2004). The following section provides a brief overview of the multimedia retrieval methods.

MULTIMEDIA QUERIES

The representation, structure, and retrieval of objects evolved in time from simple conventional data to more complex multimedia data objects such as video and audio. At the beginning, the retrieval of objects was treated as a single data item using queries based on indexing approaches and identification of the attributes that describe an object. However, this method proved to be inefficient. Retrieval of images from a multimedia database is based on similarities. Modern computer technology can be used to extract images of objects from pictures and videos, but the ability to identify these objects semantically is beyond the capabilities of current research. Because computers cannot add semantic interpretation to images, this must be added manually. Visual queries based on recognizing objects on the basis of their color, shape, or texture and other graphical characteristics are possible. Objects can be indexed using feature values and retrieved based on the similarity to the feature values of other objects. Feature values can be derived for an entire media object or just a specific part of its content. Features that can be measured include shape, color, texture, initial position, and direction and speed of motion. Features can be difficult to quantify and their identification may require complicated and lengthy techniques for extraction. The measurement value for a feature may be a representation of the shape or a histogram representing the color distribution of the object. Several of the methods for the extraction of feature values require use of advanced algorithms based on artificial intelligence (AI) techniques such as data mining (Shih & Wang, 2004; Thuraisingham, 2001), artificial neural networks and fuzzy logic (Tsai, McGarry, & Tait, 2003), and computer vision (Dunckley, 2003). These techniques are explored to enhance the feature extraction and increase the quality of multimedia retrieval and processing.

If the retrieval of images is to be based on keywords, then either the images must be manually indexed, or the objects within the images must be automatically recognized and appropriate keywords added to the index. Whether automatic or manual indexing is used, the descriptive words added by the indexer may or may not conform to the keywords employed by the user. One approach utilizes two different keyword indexing systems, conceptual keyword and scene description keyword, to retrieve images from a database. However, the developed techniques for indexing and retrieval are abundant, but there are no universally accepted techniques for feature extraction, indexing, and retrieval (Deb & Zhang, 2004).

Currently, new methods are emerging and being implemented for retrieving multimedia data objects (e.g., audio and video) based on their content. Research of content-based retrieval of multimedia information started in the early ’80s with investigations into the retrieval of static images, which in turn was based on pattern recognition research of the ’70s. Research into the retrieval of images from video has received more attention during the ’90s. The approaches to aiding the retrieval of visual images from graphics or videos can be classified into the following categories:

- Keyword, in which the content of the images is described by an indexer using keywords or a textual abstract.
- Content-based image retrieval of features that can be automatically extracted from images. The features are selected according to certain criteria, such as color, texture, or shape; by allowing the user to sketch images and then retrieve similar images from the database; by discerning the motion path of an object; and by identifying one object from its position relative to another. There are a number of commercially available systems, such as IBM's QBIC and Virage products, which are based on the retrieval of images using their non-semantic features.
- Concept-based retrieval in which semantic interpretation of the objects is added, such as identifying an object as a named person, a type of vehicle, etc. There are a number of systems in this category such as Infoscope and WebSEEk. WebSEEk is a semiautomatic system for retrieving, analyzing, categorizing, and indexing visual images from the World Wide Web.

Many algorithms have been focused on content-based image indexing, but researchers have been investigating an indexing system based on the spatiotemporal characteristics of objects that appear in multimedia applications. Spatial positions are represented by two-dimensional coordinates while temporal relationships are represented by a single dimension, time.

Other systems have been developed for automatically indexing video streams received in real time. The source of the video may be, for example, television broadcasts or security cameras. The incoming video data stream is passed to an event detection module that detects scene changes, significant audio changes, camera operations, and the motion of objects in the video. The event detection module uses this information to define the boundaries of what the authors refer to as events. The event boundaries are specific to the application; in broadcast television these may be scene changes or the appearance of captions, and in security surveillance video the beginning of a new event may be signified by a new object entering the scene. The first frame of an event is used as its key frame. The key frame, the event itself (comprising both video and audio), and its time index will be stored for a certain period of time. The key frames provide a visual index for the viewer. The view of the index displayed can be changed to show a more global view, with an index that includes every key frame. Content-based queries are often combined with text and keyword predicates to get powerful retrieval methods for image and multimedia databases.

Hypermedia video links is another system for content-based retrieval of information from a video database. Raw video data is stored in the video storage. Video clips can be extracted from the video object database using video object software. Developers of hypermedia applications can also use the video object manager to retrieve and play video data objects. The video data structure for the hypermedia multimedia authoring environment is quite straightforward. Using the video object manager, the developer of an application can extract clips manually from a video to create primitive video data objects. Attributes, such as display size and frame rate, inherited from the video from which it is extracted can be associated with the clip or altered by the developer. Complex data objects can be created using combinations of primitive and existing complex video data objects. The clips within a complex video data object are arranged in the sequence in which they will be played back. The next section introduces known multimedia management products, examples of applications using these products, and significant standards.

KNOWN MULTIMEDIA MANAGEMENT PRODUCTS

Corporations such as IBM, Virage, Silicon Graphics, ARDA, Oracle, Autonomy, and ALPHATECH are major providers of multimedia database management, communication, and content management software for corporations, media and entertainment companies, universities, and government agencies worldwide.

The IBM's CueVideo project provides technologies to automatically summarize and index videos and to make them much easier to browse. One approach is using audio stream for search and using video stream for quick browsing in a complementary manner to provide the desired video search functionality (Brown et al., 2001). CueVideo is an ongoing research project to address the challenges of large video databases.

One example of a mobile multimedia application is MobiDENK (Krosche, Baldzer, & Boll, 2004) for monument conservation. The Hermitage museum's Web site uses an IBM product, QBIC (Query By Image Content), for searching archives of world-famous art. QBIC is a system developed by IBM to explore various content-based retrieval methods. Queries in QBIC can be based on color and texture patterns, user-drawn sketches, example images, camera and object motion, or other graphical information, such as the color balance; e.g., retrieve all images with 30% blue and 10% yellow colored pixels. The QBIC database can handle both still images and video clips. In common with other systems, videos are divided into shots, using techniques similar to those described above. One frame from each clip is extracted or generated as a representative frame for the clip. In a motion sequence there may not be a single frame that is representative of the entire shot. In this case, QBIC constructs an image of the complete background from the sequence of frames and on to this superimpose images of the foreground objects.

QBIC's capabilities are available in DB2 Image Extenders, which are components of IBM's scalable, multimedia, Web-enabled DB2 Universal Database. IBM developed different multimedia retrieval applications for the DB2 Universal Database as follows:

- DB2 Audio Extender enables audio retrieval.
- DB2 Image Extender enables image retrieval. Visual features such as color and texture patterns are used for search criteria. For example, a photographic database may be queried for thumbnail images of all pictures stored in GIF format, and the name of each picture's photographer can be listed. Then, Image Extender invokes a browser.
- DB2 Video Extender enables video retrieval. Video and traditional business data may be included in a single query.

For example, Image Extender is used to store print ads, Audio Extender for broadcast ads, and Text Extender for ad scripts. The user can retrieve all the objects in a single query and then preview video ads as video storyboards using the Video Extender.

Hillsborough County uses Virage's products to create a central digital archive of government proceedings, which automates a formerly manual workflow and provides real-time access to content. Virage products include capabilities as follows:

- IDOL provides the infrastructure for capturing, feeding, and delivering rich media through the enterprise.
- VS Archive is a software solution used by enterprises to store, categorize, manage, retrieve, and distribute video, audio, and other rich media content. It includes new features such as advanced conceptual retrieval and automated hypertext linking to enterprise information found in more than 300 different data types.

- VS News Monitoring is a real-time monitoring and content management solution used by enterprises and government agencies to automatically track content for time-sensitive, strategically significant events. New features within VS News Monitoring include real-time data access and advanced concept-based retrieval. Concept-based retrieval is especially important for monitoring and searching news feeds in foreign languages, because users may work in second languages or rely on transcriptions that may contain misspelled words. With IDOL, VS News Monitoring not only proactively alerts users to broadcast news but also delivers related internal content as well as information from Web sites.
- VS Webcasting is an enterprise software solution that simplifies and streamlines the entire workflow for producing live Webcast events and on-demand or rebroadcast presentations for large audiences.

Silicon Graphics and Virage combine products to create breakthrough media management systems. For example, Virage Media Management System and the Silicon Graphics StudioCentral asset management system enable companies to automatically and intelligently catalog large libraries of videotape and multimedia content into a compact, online database. The combined products provide users with a complete media management system to find and manage their media through a simple Web browser. The next section discusses main standards for multimedia applications.

MAJOR STANDARDS

Many aspects of the multimedia information life cycle are affected by regulatory compliance (Golshani, 2004), and requirements for latest object database management development standards are proposed (Cardenas, Pon, Panayiotis, & Hsiao, 2003). MPEG-7 and MPEG-21 standards have been essential for the development of multimedia applications. The MPEG-7 standard, formally named Multimedia Content Description Interface, provides a rich set of standardized tools to describe multimedia content. Both human users and automatic systems that process audiovisual information are within the scope of MPEG-7. MPEG-7 offers a comprehensive set of audiovisual description tools (the metadata elements and their structure and relationships, which are defined by the standard in the form of descriptors and description schemes) to create descriptions (i.e., a set of instantiated description schemes and their corresponding descriptors at the user's will), which form the basis for applications enabling efficient access (search, filtering, and browsing) to multimedia content (Chang, Kikora, & Puri, 2001).

MPEG-21 has established a work plan for future standardization. Nine parts of standardization within the multimedia framework have already started.

FUTURE TRENDS

Content-based retrieval is a very active area of research despite the advances that have been made. Recently, Virage announced its participation in the Video Analysis and Content Extraction (VACE) media communication and content program sponsored by ARDA. The research project will be focused on the development of video content extraction technology. The initial focus of the VACE program is to develop automatic detection and recognition technologies from various video-related sources, including indoor and outdoor activities, unmanned air vehicles, and television news broadcasts. Over time, VACE technologies aim to provide significant improvement in indexing and retrieval, understanding, image processing, data mining, filtering and selection,

and storage and forwarding mechanisms. In addition, research is occurring on new compression techniques, such as wavelet, vector, and fractal methods, to ensure multimedia delivery with high quality.

In the future it should be possible to automatically recognize and identify objects that appear in still images and video. To achieve this, it will almost certainly require significant developments in the application of artificial intelligence techniques to multimedia database systems. These developments will lead to automatic recognition and indexing of video footage and still images and result in the development of a wide range of applications relying on the content-based retrieval of multimedia data objects.

CONCLUSION

Multimedia is becoming present everywhere: at home, business, school, hospital, road, etc. Digital media has been acknowledged as a standard data type, allowing for increased personal communications, business-to-employee, business-to-business, and business-to-consumer applications. These applications require complex and large multimedia databases, causing changes in computing and solution architectures to be maintained by organizations. In addition, technologies pertaining to communication, coding, compression, content distribution, storage, mobile computing, media servers, cryptology and watermarking, and digital media management will grow in order to address solutions for multimedia services. While creating digital media is not expensive, it is generally expensive to manage and distribute it. Browsing large multimedia databases can become complex and will demand faster and more efficient algorithms for indexing and retrieval of structures. In addition, multimedia distribution in a mobile computing environment will continue to be the center of research for next-generation multimedia systems.

REFERENCES

Bourbakis, N. (2004). Digital multimedia on demand. *IEEE Multimedia*, *11*(2), 14-15.

Brown, E. W., Srinivan, S., Coolen, A., Ponceleon, D., Cooper, J. W., & Amir, A. (2001). Towards speech as knowledge resource. *IBM Systems Journal*, *40*(4), 985-1001.

Cardenas, A. F., Pon, R. K., Panayiotis, A. M., & Hsiao, J.-T. (2003). Image stack stream viewing and access. *Journal of Visual Languages & Computing*, *14*(5), 421-441.

Chang, S.-F., Sikora, T., & Puri, A. (2001). Overview of the MPEG-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *11*(6), 688-695.

Deb, S., & Zhang, Y. (2004). An overview of content-based image retrieval techniques. *18th International Conference on Advanced Information Networking and Applications (AINA'04)* (Vol., 1, pp. 59-64).

Dunckley, L. (2003). *Multimedia databases: An object-relational approach*. London: Addison-Wesley.

Golshani, F. (2004). Multimedia information lifecycle management. *IEEE Multimedia*, *11*(2), 1.

Krosche, J., Baldzer, J., & Boll, S. (2004). MobiDENK—Mobile multimedia in monument conservation. *IEEE Multimedia*, *11*(2), 72-77.

Shih, T., & Wang, P. P. (2004). *Intelligent virtual world: Technologies and applications in distributed virtual environment*. Hackensack, NJ: World Scientific.

Shu, W., & Yu, M.-Y. (2004). Resource requirements of closed-loop video delivery. *IEEE Multimedia*, *11*(2), 24-37.

Thuraisingham, B. (2001). *Managing and mining multimedia databases*. Boca Raton, FL: CRC Press.

Tsai, C.-F., McGarry, K., & Tait, J. (2003). Using neuro-fuzzy techniques based on a two-stage mapping model for concept-based image database indexing. *IEEE Fifth International Symposium on Multimedia Software Engineering (ISMSE'03), 1,* 6-12.

Almaden (n.d.). Retrieved July 22, 2004, from http://www.almaden.ibm.com/projects/cuevideo/shtml

IBM (n.d.). Retrieved July 25, 2004, from http://wwwqbic.almaden.ibm.com

Infoscope (n.d.). Retrieved July 25, 2004, from http://www.infoscope.com

NASA Earth Observatory, http://earthob servatory. nasa.gov:8000/Laboratory/index.html, accessed July 23, 2004.

Oracle, http://www.oracle.com, accessed July 20, 2004.

TeraText, http://www.teratext.com.au, accessed July 23, 2004.

Virage, http://www.virage.com, accessed July 26, 2004.

WebSeek, http://www.ee.columbia.edu/~sfchang/research, accessed July 22, 2004.

KEY TERMS

Content-Based Retrieval: Method for automatic multimedia content features extraction.

Feature: An attribute derived from transforming the original multimedia object by using an analysis algorithm; a feature is represented by a set of numbers (also called feature vector).

Feature Extraction: Use of one or more transformations of the input features to produce more useful features.

Feature Selection: Process of identifying the most effective subset of the original features.

Indexing: Mechanism for sorting the multimedia data according to the features of interest to users to speed up retrieval of objects.

Metadata: Information about multimedia data objects, applications, processing, and delivery requirements.

Multimedia Database: A repository of different data objects such as text, graphical images, video clips, and audio.

This work was previously published in Encyclopedia of Database Technologies and Applications, edited by L. C. Rivero, J. H. Doorn, and V. E. Ferraggine, pp. 390-394, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.18
Parallel and Distributed Multimedia Databases

S. Geisler
University of Clausthal, Germany

O. Kao
University of Paderborn, Germany

INTRODUCTION

Sensing and processing of multimedia information is one of the basic traits of human beings. The development of digital technologies and applications allows the production of huge amounts of multimedia data. The rapidly decreasing prices for hardware such as digital cameras/camcorders, sound cards and the corresponding displays led to wide distribution of multimedia-capable input and output devices in all fields of the everyday life, from home entertainment to companies and educational organisations. Thus, multimedia information in terms of digital pictures, videos, and music can be created intuitively and is affordable for a broad spectrum of users.

An important question in this context is related to the archiving of the acquired information. The old-fashioned albums with pictures from holidays, children, special occasions, and so forth are replaced by photo-CDs and DVDs. Analogously, digital videos are edited, valorised by including meta-information (occasion, place, date …) and archived on DVDs. If a particular scene, image, or sound file is needed, then one can use its memory to find the corresponding medium. This type of organisation is surely not applicable to large multimedia archives, which often exist in industrial and educational sectors and where Petabytes worth of multimedia data are produced year for year. All this information has to be systematically collected, registered, stored, organised, and classified. Therefore, in many branches professional archives for such multimedia information are established, such as document management systems, digital libraries, photo and video archives used by public authorities, corporations, broadcasting and TV companies, as well as archives for satellite and surveillance photos. The scope and spread of such systems grow day by day and lead to new demands for efficient retrieval of the archived information

based on user-specific description of the sought image, video or audio.

The search for a medium similar to the given one is, due to the complexity of multimedia information, a very challenging problem and requires a number of novel mechanisms. Beside the search procedures, also methods to formulate queries, and ways to visualise the results have to be provided. Moreover, the search has to be performed efficiently in order to achieve acceptable response times for the user. Therefore, a combination of modern multimedia archives with powerful parallel and distributed architectures described in this article is mandatory for the integration of multimedia retrieval into real-world applications.

BACKGROUND

The necessity for organisation and retrieval of multimedia data led to development of a large number of prototypes and operational multimedia database management systems, which manage the multimedia data in terms of storage, annotation, and retrieval. In the early years this task was tended to by existing database management systems (DBMS) with multimedia extensions. The basis for representing and modelling multimedia data in such systems is so-called Binary Large Objects (BLOBs), which store images, video and audio sequences without any formatting and analysis done by the system. The media are saved in the current form in the database and their additional information – called meta-information – is inserted into the database tables. Typically, the file name, categories and additional key words entered by the user serve as meta-information. Once the user submits a key word about the sought media, the blocks with meta-information are searched using the existing database functions and compared with the input. In case of a key word match, the corresponding media is presented.

These extensions reflect a certain aspect of multimedia database systems, but this approach does not satisfy the requirements of multimedia archives, as the manual annotation of the media is too time-consuming and not applicable in real-world applications. Furthermore, key words are not sufficient to represent content of images or videos entirely (*An image says more than 1000 words*). Therefore, the media annotation and retrieval has to be content-based; that is, features describing the multimedia content have to be extracted automatically from the media itself and compared to the corresponding features of the sample medium. The functionality of such a multimedia database is well defined by Khoshafian and Baker, (1996):

"A multimedia database system consists of a high performance DBMS and a database with a large storage capacity, which supports and manages, in addition to alphanumerical data types, multimedia objects with respect to storage, querying, and searching."

The DBMS is already provided by traditional databases and therefore will not be discussed in the following sections. Instead, the focus is set on the mechanisms for multimedia retrieval and high-performance implementation.

MULTIMEDIA RETRIEVAL

The content-based annotation of multimedia data requires the integration of additional information, which can be classified into the following categories:

- Technical information describes details of the recording, conversion, and storage. Examples: filename, resolution, compression, frame rate.
- Extracted attributes are features that are deduced by analysing the content of the media directly. Examples: average colour or colour histograms of an image, camera motion in videos, pitch in audio files.

Table 1. Prominent examples for image, video and audio databases.

• Image databases
  o Qbic (Flickner, Sawhney et al., 1995)
  o Photobook (Pentland, Picard & S claroff, 1994)
  o Surfimage (Nastar, Mitschke et al., 1998)
• Audio databases
  o VARIATIONS (Dunn & Mayer, 1999)
  o MUSART (Birmingham, Dannenberg et al., 2001)
• Video database s
  o VideoQ (Chang, Chen et al., 1998)
  o Virage Video Engine (Hampapur, Gupta et al., 1997)
  o CueVideo (Ponceleon, Srinivasan et al., 1998)

- Knowledge-based information links the objects, people, scenarios, and so forth detected in the media to entities in the real world.
- World-oriented information encompasses information on the producer of the media, the date and location, language, and so forth.

Technical and world-oriented information can be modelled with traditional data structures. The knowledge-based information assumes semantic analyses of the media, which is nowadays still not possible in general. However, many recent research efforts in this direction promise the applicability of semantic information in the future (Zhao & Grosky, 2002).

Most of the existing multimedia retrieval systems are specialised to work on media of a limited domain, for example news (Christel & Hauptman, 2002; Yang & Chairsorn, 2003), American football (Li & Sezan, 2002), or integrate general retrieval algorithms like face, speech or character recognition. They use features extracted from the media content to annotate and retrieve the multimedia objects, which are usually related to the colour, edge, texture, layout properties in case of images or consider object motion in case of videos or specific tone sequences for audios. Table 1 gives an overview of several prominent, specialised systems, which introduced main research retrieval concepts to the scientific community. Meanwhile, many of these systems became a part of commercial products or a part of a general multimedia archive. A survey is provided in Venters and Cooper (2000).

In the following, the retrieval workflow for multimedia data will be depicted by considering images as an example. The user has a specific image in mind and starts a query for this specific image or for similar samples. The user can for example browse the data set or give a suitable key word. However, for content-based similarity search, sophisticated interfaces are necessary:

- Query by pictorial example: the user supplies the system with a complete sample image, which is similar to the sought one.
- Query by painting: the user sketches the looked-for image with a few drawing tools (Rajendran & Chang, 2000).
- Selection from standards: lists of sample instances – called standards – can be offered for individual features.
- Image montage: the image is composed of single parts similar to a mosaic.

All approaches have individual advantages as well as shortcomings; thus a suitable selection depends on the domain. However, the query by pictorial example is one of the most often applied methods since it provides the greatest degree of flexibility.

In the next step the query image is compared to all archived images in the database based on the extracted features. Each feature emphasises one or more aspects of the image, usually related to colour, texture and layout. These features are extracted off-line (at creation time) for the archived images and online (at query time) for the given sample image. Thus, after the analytical phase, the images are represented by consistent feature vectors, which can be directly compared using similarity metrics or functions. For this purpose well-known metrics such as the (weighted) Euclidian Distance or specially developed and adapted metrics like Earth Mover Distance can be applied (Rubner & Tomasi, 2000).

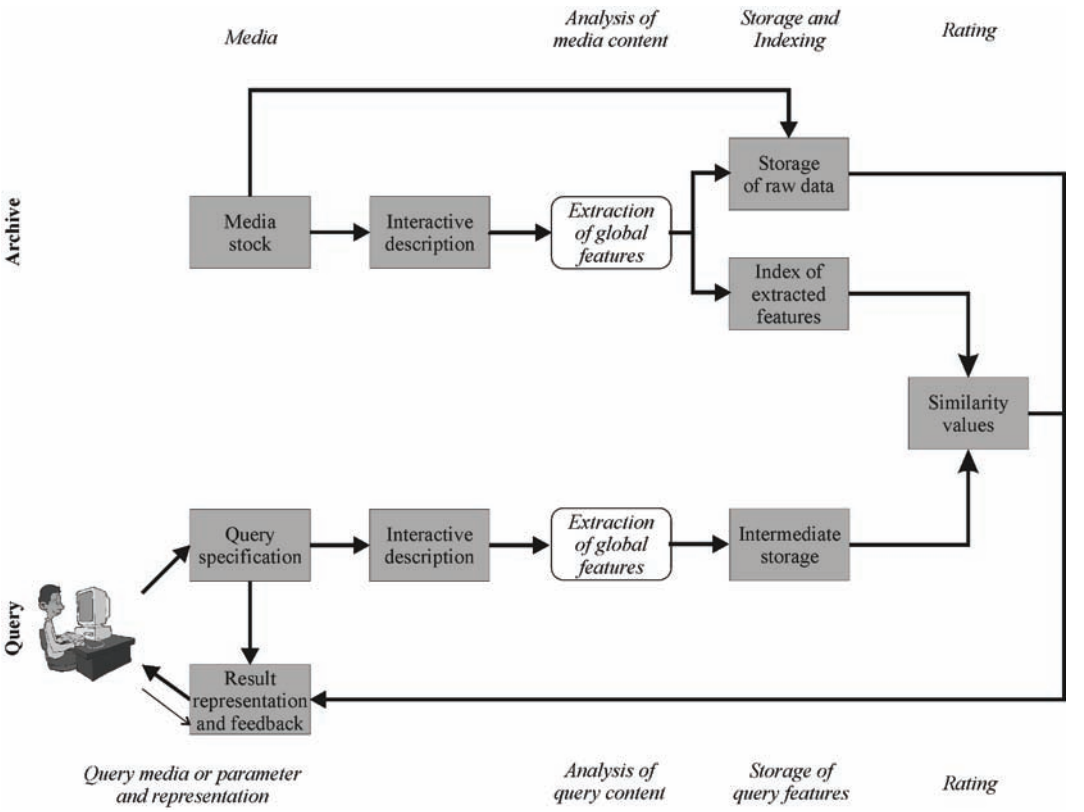
The result of this comparison is a similarity value for the query and the analysed image. The process is repeated for all n images in the database, resulting in a similarity ranking. The first k entries, k being a user-defined constant, represent the k best hits, whose raw data are then displayed. Figure 1 describes the global workflow.

The comparison process can be accelerated by using index structures. These contain a-priori extracted features and are organised in a way that the comparisons can be focused to a certain area around the query. A novel approach and an overview of often-applied index structures are provided in Tuncel and Ferhatosmanoglu (2002). A survey of specialised retrieval techniques is provided for example in Rui (1999).

HIGH-PERFORMANCE COMPUTING FOR MULTIMEDIA RETRIEVAL

The database architecture – software and hardware – is decisive for the efficiency and thus for

Figure 1. General retrieval workflow



the usability of a multimedia archive. Because of their high storage and computing requirements, multimedia databases belong to those applications which rapidly hit the limits of the existing technology.

The widespread client/server architectures are – in their usual form – not fit for multimedia database implementations. Firstly, a centralised organisation of a media server requires immense storage and computation resources. With a growing number of user queries and data to be organised, such a centralised system will quickly reach the borders of its capabilities and the quality of service is no longer fully sustainable.

One possible solution for this problem is offered by distributed and parallel architectures, where multiple processing elements (PEs) work cooperatively on an efficient solution of a large problem. The data and the programs are spread over several nodes, so that the processing is accelerated (parallel processing) or the path to the user is shortened (distributed processing, e.g., video-on-demand servers).

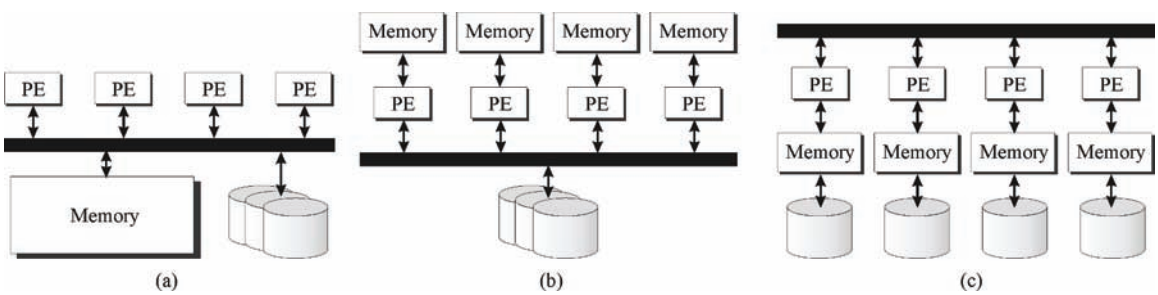
There are many possible ways for the organisation of such architecture; the well-known are shared everything, shared disk, and shared everything architectures shown in Figure 2.

Due to the shared utilisation of the hard disks, buses, and memory in case of shared disk and shared everything architecture respectively, the data throughput in these architectures is rather low and the response time long. Experimental measurements executed in Bretschneider, Geisler and Kao (2002) showed that shared nothing architectures – in particular cluster-based systems – with an appropriate data distribution across the multiple nodes are most suitable for the realisation of an efficient multimedia retrieval. A node is defined as an entity consisting of a PE, a memory, a storage resource, and a network adapter.

The database running on this architecture can be denoted as a parallel and distributed database: The data are distributed over all nodes, but logically combined in a single entity, making it look like a single database system from the outside. The only difference noticeable to the user is the performance improvement of the system, which results from utilising the parallelism in the computer network. Further characteristics concern the optimisation of queries, controlling parallelism, recovery, integrity, and security. All these aspects are already well defined for conventional database systems, and have been analysed in detail. However, the possibility of a geographically separated data distribution introduces new constraints and requires additional communication and synchronisation mechanisms. The problems grow more complex when heterogeneous computer architectures, long processing times per data set – as in multimedia applications –, hardware and software failures, and so forth have to be considered.

As already noted, the data distribution is a crucial efficiency aspect and has to fulfil numerous, sometimes even conflicting requirements: The

Figure 2. Classification of parallel architectures: a) shared everything; b) shared disk; c) shared nothing



data needed by an operation should – if possible – all be on one node (*data locality*). On the other hand, as many operations as possible should be processed in parallel, that is, the data should be distributed evenly among all available nodes. A drawback of the broad distribution is given by the time-consuming data transfer between the individual nodes, which affects the performance significantly. Thus a general distribution is not applicable; hence it has to be tailored for the current application. One possible solution will be presented in the following by considering the cluster-based multimedia database Cairo as an example (Kao & Stapel, 2001).

The general Cairo architecture shown in Figure 3 consists of:

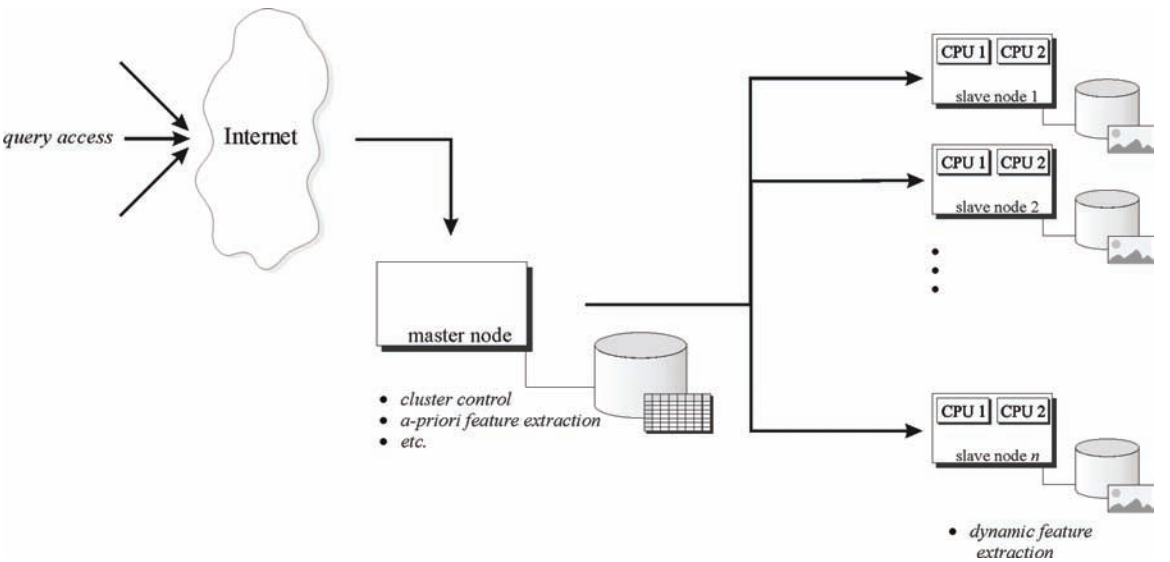
- Query stations host the Web-based user interfaces for the access to the database and for the visualisation of the retrieval results.
- Master node controls the cluster, receives the query requests and broadcasts the algorithms, search parameters, the sample media, features, and so forth to the compute nodes. Furthermore, it unifies the intermediate results of the compute nodes and produces the final list with the best hits.
- Compute nodes perform the image processing and comparisons. Each of these nodes contains at least one partition and executes all operations with the data stored on the local devices. The computed intermediate results are sent to the master node.

The distribution of the media stock across the individual cluster nodes is created according to the following requirements:

- Similar storage sizes of the partitions and thus an even distribution of the media across the individual nodes,
- Computation reduction for the retrieval, and
- Minimising the communication between the cluster nodes.

Cairo uses a content-independent, size-based strategy for the available media stock that creates a set of partitions $P = P_1 \cup P_2 \cup \ldots \cup P_n$ where $P_i \cap P_j = \varnothing$ and $size(P_i) \approx size(P_j)$ for all i,j, $i \neq j$. The

Figure 3. Schematic of the Cairo cluster architecture



current image under investigation is assigned to the node with the least storage space used. This process is repeated until all images are distributed to the partitions.

The processing of a partition P_i is executed per media instance. The individual operations are independent of one another, so the order of execution is irrelevant. This initial partitioning makes it possible for all nodes to have uniform processing times, assuming a homogenous, dedicated execution platform. The management overhead depends on the operator applied and the structure of the partial results. However, this time is usually neglectable, compared to the media processing times.

The advantage of a static distribution is that expensive computations and data transfers are not necessary during runtime and the partitioning can be manually optimised and adapted to a given application. On the other hand, short-term reactions to variable workloads among the nodes are not possible. A dynamic distribution of the data is done during runtime and considers the current workload of the nodes, as well as the number and the structure of pending queries. Idle times can be minimised by continuously re-distributing the data, and increasing the total throughput. Unfortunately, the long communication time for multimedia data often eliminates all performance advantages gained through the parallel processing.

After the partitions are created, these are distributed evenly over the cluster nodes, allowing parallel retrieval by executing the same operations on all nodes with the local media subset. The implementation is based on the following components:

- Transaction manager encompasses the analysis of the transformations to be executed and the order of the operations.
- Distribution manager receives a list of algorithms to be executed and a set of media instances as input. Then the programs for the analysis and comparison are compiled and sent to the computing managers.
- Computing manager controls the execution of the extraction algorithms with the local data. The process runs on each cluster node and supervises the communication with the master node.
- Result manager: The data partitioning in disjoint subsets results in a sub-ranking per node that need to be unified by the result manager.
- Update manager realises the media insertion in the database. The raw data are transformed in a uniform format and tagged with a unique identifier. All procedures for feature extraction are then applied. Finally, the cluster node, on whose hard disk the raw data are to be stored, is determined.

The described organisation was successfully implemented and evaluated in real-world scenarios. The minimised communication between the nodes allows a nearly full utilisation of the available compute resources and leads to a linear speedup; that is, the response time is shortened by a factor equal to the number of included nodes.

Other performance relevant parts of a multimedia database system are still the subject of research. Examples are given by the disk scheduler for multimedia storage (Huang & Huang, 2004) and scheduling for broadcasting videos and audios (Yoshihisa & Tsukamoto, 2003).

FUTURE TRENDS

Despite the immense development of multimedia in recent years, the efficient organisation and retrieval of multimedia data still remains a large scientific challenge. The growing number of multimedia data and new applications fields are the driving forces.

The consideration of semantic information is necessary to realise a human-like retrieval

approach for images, video and audio and to integrate multimedia components in the semantic Web for ubiquitous access. Furthermore, the current restriction of the most databases to a single media type will be eliminated by methods for multimodal retrieval. For example, the editor of a newscast needs access to videos from news agencies, images from photographs, audios files from radio and text from newspaper articles. Merging the information into a single database requires methods to classify all these media instances and measure the content-based similarity between them. Speech recognition for example can be used to compare audio or video files with textual information. Multimodal access accommodates different user profiles or device capabilities and can help to receive improved query results.

Not only different media types but also media data from different providers have to be merged to create a large distributed multimedia warehouse or a multimedia grid. MPEG-21 is designed for the content description and digital rights management. Load balancing, data distribution and quality of service considerations are necessary to achieve high performance and consumer satisfaction in a very heterogeneous environment. Finally, the multimedia information has to be accessible for mobile devices using standards such as UMTS.

CONCLUSION

Traditional databases offer a search for certain media instance based on describing key words. The main invention of multimedia databases is the possibility to search directly on the media content. However, this time-consuming process has to be performed in parallel to achieve acceptable response times. The cluster architecture scales well even for a high number of processing elements, but if and only if the underlying data partitioning strategy is suitable for the given application.

Further research is needed to give the user feedback on how to improve the query and how to supply the system with feedback about the relevance of delivered results. A further aspect is the improvement of the database design, especially data distribution in parallel environments. Furthermore, advanced methods for semantic content description have to be exploited.

REFERENCES

Bainbridge, D. et al. (1999). Towards a digital library of popular music. *Proceedings of the 4th ACM Conference on Digital Libraries* (pp. 161-169).

Birmingham, W.P. et al. (2001). MUSART: Music retrieval via aural queries. *Int. Symposium on Music Information Retrieval (ISMIR).*

Bretschneider, T., Geisler, S., & Kao, O. (2002). Simulation-based assessment of parallel architectures for image databases. *Proceedings of the Conference on Parallel Computing* (pp. 401-408). Imperial College Press.

Chang, S.-F. et al. (1998). A fully automated content based video search engine supporting spatio-temporal queries. *IEEE Trans. CSVT, 8*(5), 602-615.

Christel, M., Ng, H., & Wactlar, A.H. (2002). Collages as dynamic summaries for news video. *Proceedings of ACM Multimedia '02* (pp. 561-569).

Dunn, J.W., & Mayer, C.A. (1999). VARIATIONS: A digital music library system at Indiana University. *Proceedings of the 4th ACM Conference on Digital Libraries* (pp. 12-19).

Flickner, M. et al. (1995). Query by image content the QBIC system. *IEEE Computer Magazine, 28*(9), 23-32.

Hampapur, A. et al. (1997). Virage Video Engine. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases, 3022,* 188-197.

Huang, Y.-F., & Huang, J.-M. (2004). Disk scheduling on multimedia storage servers. *IEEE Transactions on Computers, 53*(1), 77-82.

Kao, O., & Stapel, S. (2001). Case study: Cairo – a distributed image retrieval system for cluster architectures. In T.K. Shih (Ed.), *Distributed multimedia databases: Techniques and applications* (pp. 291-303). Hershey, PA: Idea Group Publishing.

Khoshafian, S., & Baker, A.B. (1996). *Multimedia and imaging databases.* Morgan Kaufmann Publishers.

Li, B., & Sezan, M.I. (2002). Event detection and summarization in American football broadcast video. *Proceedings SPIE Storage and Retrieval for Media Databases, 4676,* 202-213.

Nastar, C., Mitschke, M., Meilhac, C., & Boujemaa, N. (1998). Surfimage: A flexible content-based image retrieval system. *ACM Multimedia'98 Conference Proceedings* (pp. 339-344).

Pentland, A., Picard, R., & Sclaroff, S. (1994). Photobook: Tools for content-based manipulation of image databases. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases II, 2185,* 34-47.

Ponceleon, D., Srinivasan, S., Amir, A., Petkovic, D., & Diklic, D. (1998). Key to effective video retrieval: Effective cataloging and browsing. *ACM Multimedia'98 Conference Proceedings* (pp. 99-107).

Rubner, Y., Tomasi, C., & Guibas, L.J. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision, 40*(2), 99-121.

Rui, Y., Huang, T., & Chang, S. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation, 10*(4), 39-62.

Tuncel, E., Ferhatosmanoglu, H., & Rose, K. (2002). VQ-index: An index structure for similarity searching in multimedia databases. *ACM Multimedia 2002* (pp. 543-552).

Venters, C.C., & Cooper, M. (2000). A review of content-based image retrieval systems. *Technical Report jtap-054.* University of Manchester.

Yang, H., Chaisorn, L., Zhao, Y., Neo, S.-Y., & Chua, T.-S. (2003). VideoQA: Question answering on news video. *ACM Multimedia 2003* (pp. 632-641).

Yoshihisa, T., Tsukamoto, M., & Nishio, S. (2003). Scheduling methods for broadcasting multiple continuous media data. *ACM MMDB '03* (pp. 40-47).

Zhao, R., & Grosky, W.I. (2002). Bridging the semantic gap in image retrieval. In *Distributed multimedia databases: Techniques & applications* (pp. 14-36). Hershey, PA: Idea Group Publishing.

Zhao, R., & Grosky, W.I. (2002). Narrowing the semantic gap–Improved text-based Web document retrieval using visual features. *IEEE Transactions on Multimedia, 4*(2), 189-200.

KEY TERMS

A-Priori Feature Extraction: Analysis and description of the media content at the time of insertion in the database. The gained information is stored in a database and enables content-based retrieval of the corresponding media object without actually accessing the latter.

Cluster: Parallel architecture that contains multiple "standard" computers connected via a high performance network that work together to solve the problem.

Dynamic Feature Extraction: Analysis and description of the media content at the time of querying the database. The information is computed on demand and discarded after the query was processed.

Feature Vector: Data that describe the content of the corresponding multimedia object. The elements of the feature vector represent the extracted descriptive information with respect to the utilised analysis.

Load Balancing: Techniques to distribute the tasks over the single processors in parallel systems in a way that idle time is minimized.

Multimedia Database (after Khoshafian & Baker): A multimedia database system consists of a high performance database management system and a database with a large storage capacity, which supports and manages, in addition to alphanumerical data types, multimedia objects with respect to storage, querying, and searching (Khoshafian & Baker, 1996).

Quality of Service (QoS): The collective term for all demands on the recording and the replaying procedures, which refer to generating and maintaining a continuous data stream.

Query by Example/Sketch/Humming: Methods to formulate queries in multimedia databases. The user provides an example media file and the result of the database system is a set of similar media files. If no example is available the user can draw a sketch or hum a melody.

Retrieval: Accessing stored information from the database.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 4, edited by M. Khosrow-Pour, pp. 2265-2271, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.19
Multimedia Information Filtering

Minaz J. Parmar
Brunel University, UK

Marios C. Angelides
Brunel University, UK

INTRODUCTION

In the film *Minority Report* (20th Century Fox, 2002), which is set in the near future, there is a scene where a man walks into a department store and is confronted by a holographic shop assistant. The holographic shop assistant recognises the potential customer by iris-recognition technology. The holographic assistant then welcomes the man by his name and starts to inform him of offers and items that he would be interested in based on his past purchases and what other shoppers who have similar tastes have purchased. This example of future personalised shopping assistants that can help a customer find shopping goods is not too far away from becoming reality in some form or another.

Malone, Grant, Turbak, Brobst, and Cohen (1987) introduced three paradigms for information selection, *cognitive*, *economic*, and *social*, based on their work with a system they called the Information Lens. Their definition of cognitive filtering, the approach actually implemented by the Information Lens, is equivalent to the "content filter" defined earlier by Denning, and this approach is now commonly referred to as "content-based" filtering. Their most important contribution was to introduce an alternative approach that they called social (now also more commonly called collaborative) filtering. In social filtering, the representation of a document is based on annotations to that document made by prior readers of the document.

In the 1990s much work was done on collaborative filtering (CF). There were three systems that were considered to be the quintessential recommender systems. The Grouplens project (Miller, Albert, Lam, Konstan, & Riedl, 2003) initially was used for filtering items from the Usenet news domain. This later became the basis of Movielens. The Bellcore Video recommender system (Hill, Stead, Rosenstein, & Furnas, 1995), which rec-

ommended video films to users based on what they had rented before, and Ringo (Shardanand & Maes, 1995), which later was published on the Web and marketed as Firefly, used social filtering to recommend movies and music.

BACKGROUND

Filtering multimedia content is an extensive process that involves extracting and modeling semantic and structural information about the content as well as metadata (Angelides, 2003). The problem with multimedia content is that the information presented in any document is multimodal by definition. Attributes of different types of media vary considerably in the way the format of the content is stored and perceived. There is no direct way of correlating the semantic content of a video stream with that of an audio stream unless it is done manually. A content model of the spatial and temporal characteristics of the objects can be used to define the actions the objects take part in. This content model can then be filtered against a user profile to allow granular filtering of the content, allowing for effective ranking and relevancy of the documents.

Filtering has mainly been investigated in the domain of text documents. The user's preferences are used as keywords, which are used by the filters as criteria for separating the textual documents into relevant and irrelevant content. The more positive keywords contained in a document, the more relevant the document becomes. Techniques such as latent semantic indexing have found ways of interpreting the meaning of a word in different contexts to allow accurate filtering of documents using different syntax, but allow the same semantics to be recognised and understood.

Text documents adhere to the standards of the language they are written in. Trying to do the same for AV data streams, you are faced with the problem of identifying the terms in the content itself. The terms are represented as a series of objects that appear in the content, for example, a face in an image file. These terms cannot be directly related to the objects as there is no method of comparison, or if there is, it is complex to unlock. The title of the document and some information might be provided in the file description, but the actions and spatial and temporal characteristics of the objects will not be described to a sufficient level for effective analysis of relevancy.

MAIN THRUST OF ARTICLE

Information-filtering techniques have been applied to several areas including American football (Babaguchi, Kawai, & Kitahashi, 2001), digital television (Marusic & Leban, 2002), Web applications (Kohrs & Merialdo, 2000), and ubiquitous and pervasive device applications (Tseng, Lin, & Smith, 2002).

Filtering multimedia information requires different approaches depending on the domain and use of the information. There are two main types of multimedia information filtering: collaborative and content based. If the user wants a subjective analysis of content in order to find a recommendation based on their individual preference, then they use collaborative filtering, also known as social or community-based filtering. If, on the other hand, they require an objective decision to filter information from a data stream based on their information needs, then they use content-based filtering.

All of the above systems use either collaborative or content-based filtering or a combination of both (hybrid) as the techniques for recommending predictions on candidate objects. There are existing information-filtering models outside these classic techniques such as temperament-based filtering (Lin & McLeod, 2002), which looks at predicting items of interest based on temperament theory. It works on the same principle as social filtering. Unlike social filtering, the users are grouped on temperaments of the users and not on similar item selection.

Content-Based Filtering

Content-based filtering is suited to environments where the user requires items that have certain content features that they prefer. Collaborative filtering is unsuitable in this environment because it offers opinions on items that reflect preferences for that user instead of providing filtering criteria that tries to disseminate preferred content from a data stream based on a user's preference. Personalised video summaries are the perfect domain to use content-based filtering. The reason for this is that a user will be interested in certain content only within any video data stream. For example, when watching a football game, the user may only be interested in goals and free kicks. Therefore, users can state what content features and other viewing requirements they prefer and then filter the footage against those requirements.

The content-based approach to information filtering has its roots in the information retrieval (IR) community and employs many of its techniques. The most prominent example of content-based filtering is the filtering of text objects (e.g., mail messages, newsgroup postings, or Web pages) based on the words contained in their textual representations. Each object, here, text documents, is assigned one or more index terms selected to represent the best meaning of the document. These index terms are searched to locate documents related to queries expressed in words taken from the index language. The assumption underlying this form of filtering is that the "meaning" of objects and queries can be captured in specific words or phrases. A content-based filtering system selects items based on the correlation between the content of the items and the user's preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences (van Meteren & Someren, 2000).

The main problem with content-based filtering is that it does not perform well in domains were the content of items is minimal and the content cannot be analysed easily by automatic methods of content-based retrieval (e.g., ideas and opinions). Users with eclectic tastes or who make ad hoc choices are given bad recommendations based on previous choices. For example, Dad, who usually buys classic rock CDs for himself, purchases a So Solid Crew album for his 12-year-old son. He may start getting recommendations for hardcore garage dance anthems every time he logs in. CF does not suffer this problem as it will rank on other users' recommendations of similar choices. Comparative studies have shown that collaborative-filtering recommender systems on the whole outperform content-based filtering.

Collaborative Filtering

A purely content-based approach to information filtering is limited by the process of content analysis. In some domains, until recently, the items were not amenable to any useful feature extraction with content-based filtering (such as movies, music, restaurants). Even for text documents, the representations capture only certain aspects of the content, and there are many others that would influence a user's experience, for example, in how far it matches the user's taste (Balabanovic, 2000).

Collaborative filtering is an approach to overcome this limitation. The basic concept of CF is to automate social processes such as "word of mouth." In everyday life, people rely on the recommendations from other people either by word of mouth, recommendation letters, and movie and book reviews printed in newspapers. Collaborative filtering systems assist and augment this process and help people in making decisions.

There are two main drawbacks to using collaborative filtering: the sparsity of large user-item databases and the first-rater problem (Rashid et al., 2002). Sparsity is a condition when not enough ratings are available due to an insufficient amount of users or too few ratings per user. An example of sparsity is a travel agent Web site, which has

tens of thousands of locations. Any user on the system will not have traveled to even 1% of the locations (possibly thousands of locations). If a nearest-neighbour algorithm is used, the accuracy of any recommendation will be poor as a sufficient amount of peers will not be available in the user-item database. The first-rater problem is exhibited when a new user is introduced that has not enough ratings. If no ratings have been given for an item or a new user has not expressed enough opinions, choices, or ratings, no predictions can be made due to the insufficient data available or bad recommendations will be made. In contrast, content-based schemes are less sensible to sparsity of ratings and the first-rater problem since the performance for one user relies exclusively on his or user profile and not on the number of users in the system.

Hybrid Filtering

Both content-based and collaborative filtering have disadvantages that decrease the performance and accuracy of the systems that implement them. If these methods are combined, then the drawbacks of one technique can be counteracted by the techniques of the other, and vice versa. There have been various implementations such as the following.

- By making collaborative recommendations, we can use others' experiences as a basis rather than the incomplete and imprecise content-analysis methods.
- By making content recommendations, we can deal with items unseen by others.
- By using the content profile, we make good recommendations to users even if there are no other users similar to them. We can also filter out items.
- We can make collaborative recommendations between users who have not rated any of the same items (as long as they have rated similar items).
- By utilizing group feedback, we potentially require fewer cycles to achieve the same level of personalisation.

User Profiles

In information filtering, a user's needs are translated into preference data files called user profiles. These profiles represent the users' long-term information needs (Kuflik & Shoval, 2000). The main drawbacks of using user profiles are creating a user profile for multiple domains and updating a user profile incrementally. The user profile can be populated by one or more of the following.

- *Explicit profiling*: This type of profiling allows users to let the Web site know directly what they want. Each user entering the site will fill out some kind of online form that asks questions related to a user's preferences (Eirinaki & Vazirgiannis, 2003). The problem with this method is the static nature of the user profile once it has been created. The stored preferences in the user profile cannot take into account the changing user's preferences.
- *Implicit user profiles*: This type of user profiles is created dynamically by tracking the user's behaviour pattern through automatic extraction of user preferences using some sort of software agent, for example, intelligent agents, Web crawlers, and so forth (Eirinaki & Vazirgiannis, 2003). All these usage statistics are correlated into a usage history that is an accurate interaction between the user and the system. This usage history is then analysed to produce a user profile that portrays the user's interests. The user profile can be updated every time the user starts a new session, making implicitly made profiles dynamic. The downside of this method is that the user initially will have to navigate and explore the site before enough data can be generated to produce an accurate profile.

Table 1. Multimedia information filtering

	Description	Techniques Used	Advantages	Disadvantages	Future R&D
Information Filtering	filtering a dynamic information space using relatively stable user requirements	SDI systems recommender systems	allows user to constantly receive content they are interested in with minimal user effort	does not support ad hoc queries that are dynamic compared to the information space they are searching (information retrieval)	all of the below
Content-Based Filtering	filtering content from a data stream based on extracting content features that have been expressed in a content-based user profile	vector space model probabilistic/inference models latent semantic indexing	objective analysis of large and/or complicated (e.g., multimedia) sources of digital material without much user involvement	1. content dependent 2. hard to introduce serendipitous recommendations as approach suffers from "tunnel vision" effect	extracting semantics from the structure of the content automatically without human intervention
Collaborative Filtering	filtering items based on similarities between target user's collaborative profile and peer users/ group	same as above	1. content independent 2. proves more accurate than content-based filtering for most domains of use enables introduction of serendipitous choices	1. sparsity: poor prediction capabilities when new item is introduced to database due to lack of ratings 2. new user: poor recommendations made to new users until they have enough ratings in their profiles for accurate comparison to other users	solving the sparsity and new-user problem finding other types of ratings schemes that do not use comparisons between users tastes (e.g., filtering using users temperament)
Hybrid Filtering	combines two or more filtering techniques	simple or rule based stereotype collaborative content based	to reduce weak points and promote strong points of each of the techniques used	weak points can outweigh strong points if the hybrid is created naively	using hybrid systems in domains where using one technique presents a large disadvantage/problem
User Profiles	log file containing user's preferences for consumption of content	content-based profiles collaborative profiles	1. user does not need to state preferences each and every time they use the system 2. user can maintain and update preferences with minimal effort compared to ad hoc methods	needs frequent updating or user preferences become stagnant	user profiles that are as ubiquitous and pervasive as the devices/systems that use them standardisation

continued on following page

Table 1. continued

Explicit User Profiles	user manually creates user profile by means of a questionnaire	questionnaires ratings	preference information gathered is usually of high quality	requires a lot of effort from user to update	collecting new user preferences that reduces user effort
Implicit User Profiles	system generates user profile from usage history of interactions between user and content	machine learning algorithms	minimal user effort required easily updateable by automatic methods	initially requires a large amount of interaction between user and content before an accurate profile is created	new machine learning algorithms for better accuracy when creating implicit user profiles
Hybrid User Profiles	combination of user profile techniques used to create a profile	explicit/implicit user profiles	to reduce weak points and promote strong points of each of the techniques used	N/A	finding effective strategies for deployment and use of hybrid profiles

- *Hybrid of implicit and explicit profiling*: The drawbacks of explicit and implicit profiling can be overcome by combining both methods into a hybrid. This allows the strong points of one technique to counteract the shortcomings of the other and vice versa. The hybrid method works by collecting the initial data explicitly using an online form. This explicitly created data is then updated by the implicit tracking method as the user navigates around the site. This is a more efficient method over both pure methods. In some instances, this hybrid method is reversed and the implicit tracking methods are used initially to produce a profile.
- *Stereotype profiling*: This can be achieved by data mining and analysis of usage histories over a period of visits. This provides accurate profiling for existing users with legacy data that is accurate. The disadvantage of this method is that it suffers from the same static nature as explicit profiling as the profile is created from archive data that might be obsolete, and therefore some updating might be necessary. The predefined user stereotype is a content-based user profile that has been created for a virtual user or group of users who have common usage and filtering requirements for consumption of certain material. The stereotyped profile will contain additional information about the stereotyped user such as demographic and social attributes. This additional information is then used to place new users to stereotyped profiles that match similar demographic and social traits. The new user without the need of any implicit or explicit tracking automatically inherits preference information.

FUTURE TRENDS

Content-based filtering in multimedia information filtering has one innate problem that researchers

are trying to solve: How can we extract semantics automatically from structural content of the model? In collaborative filtering, the age-old problem of sparsity and the new-user problem are still the biggest hindrances to using this method of filtering. Sparsity is being solved presently by hybrid systems, and it appears that this will be favoured way of dealing with sparsity (Lin & McLeod, 2002). The most promising solutions appear to be collaboratively filtering, standardised content-based profiles, which allow flexibility for systems to use either pure content-based or collaborative filtering, or a hybrid of both interchangeably.

Current work on user profiles focuses on improving creation techniques such as improved machine learning algorithms that create implicit user profiles more rapidly so that they can be more reliable and accurate in a shorter amount of time. For explicit user profiling, there is the work on selecting items that increase the usefulness of initial ratings that we have already discussed. The main way forward here, though, appears to be hybrid user profiles that are initially explicitly created and then implicitly updated.

With the advent of digital television and broadband, consumers will be faced with a deluge of multimedia content available to them at home and at work. What they will require are autonomous, intelligent filtering agents and automated recommender systems that actively filter information from multiple content sources. These personalisation systems can then collaborate to produce ranked lists of recommendations for all purposes of information the user might require. The key to this kind of service is not in the implementation of these systems or the way they are designed, but rather on a standard metadata language that will allow systems to communicate without proprietary restrictions and aid in end-user transparency in the recommendation process.

CONCLUSION

In the coming years, as nearly all communication and information devices become digital, we will see the development of systems that will be able not only to recommend items of interest to us, but will be able to make minor decisions for us based on our everyday needs such as ordering basic shopping groceries or subscribing to entertainment services on an ad hoc basis. What is required is a model of the user that describes the user's preferences for a multitude of characteristics that define the user's information needs. This model can then be used to filter data and recommend information based on this complete view of the user's needs. This has been done for many years with text files using techniques such as content-based and collaborative filtering, but has always been a problem with multimedia as the content is diverse in terms of storage, analysis techniques, and presentation. In recent years, classical techniques used for text filtering have been transferred and used in the area of multimedia information filtering. New developments such as hybrid filtering and improved metadata languages have made filtering multimedia documents more reliable and closer to becoming a real-world application.

REFERENCES

20th Century Fox Pictures. (2002). *Minority report* [Motion picture]. 20th Century Fox Pictures.

Angelides, M. C. (2003). Guest editor's introduction: Multimedia content modelling and personalization. *IEEE Multimedia, 10*(4), 12-15.

Babaguchi, N., Kawai, Y., & Kitahashi, T. (2001). Generation of personalised abstract of sports and video. *IEEE Expo 2001,* 800-803.

Balabanovic, M. (2000). An adaptive Web page recommendation service. *First International Conference on Autonomous Agents,* 378-385.

Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for Web personalization. *ACM Transactions on Internet Technology (TOIT), 3*(1), 1-27.

Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 194-201.

Kohrs, A., & Merialdo, B. (2000). Using category-based collaborative filtering in the active Web museum. *IEEE Expo 2000.*

Kuflik, T., & Shoval, P. (2000). Generation of user profiles for information filtering: Research agenda. *Proceedings of 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 313-315.

Lin, C., & McLeod, D. (2002). Exploiting and learning human temperaments for customized information recommendation. *Internet and Multimedia Systems and Applications*, 218-223.

Malone, T. W., Grant, K. R., Turbak, F. A., Brobst, S. A., & Cohen, M. D. (1987). Intelligent information sharing systems. *Communications of the ACM, 30*(5), 390-402.

Marusic, B., & Leban, M. (2002). The myTV system: A digital interactive television platform implementation. *IEEE Expo 2002.*

Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., & Riedl, J. (2003). MovieLens unplugged: Experiences with an occasionally connected recommender system. *Proceedings of ACM 2003 International Conference on Intelligent User Interfaces (IUI'03).*

Rashid, M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., et al. (2002). Getting to know you: Learning new user preferences in recommender systems.

Shardanand, U., & Maes, P. (1995). *Social information filtering: Algorithms for automating "word of mouth."* Proceedings of the CHI-95 Conference, Denver, CO.

Tseng, B. L., Lin, C.-Y., & Smith, J. R. (2002). Video summarization and personalization for pervasive mobile devices. *SPIE* (Vol. 4676). San Jose.

van Meteren, R., & Someren, M. (2000). *Using content-based filtering for recommendation.* Retrieved from http://citeseer.nj.nec.com/499652.html

Wyle, M. F., & Frei, H. P. (1989). Retrieving highly dynamic, widely distributed information. In N. J. Belkin & C. J. van Rijsbergen (Eds.), *Proceedings of the 12th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval* (pp. 108-115). ACM.

KEY TERMS

Collaborative Filtering: Aims at exploiting preference behaviour and qualities of other persons in speculating about the preferences of a particular individual.

Content-Based Filtering: Organizes information based on properties of the object of preference and/or the carrier of information.

Hybrid Filtering: A combination of filtering techniques in which the disadvantages of one type of filtering is counteracted by the advantages of another.

Information Filtering: Filtering information from a dynamic information space based on a user's long-term information needs.

Recommendation: A filtered list of alternatives (items of interest) that support a decision-making process.

Recommender Systems: Assist and augment the transfer of recommendations between members of a community.

User profile : a data log representing a model of a user that can be used to ascertain behaviour and taste preferences.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 4, edited by M. Khosrow-Pour, pp. 2063-2068, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.20
Multimedia Information Retrieval at a Crossroad: Review and Outlook

Qing Li
City University of Hong Kong, Hong Kong

Jun Yang
Carnegie Mellon University, USA

Yueting Zhuang
Zhejiang University, China

INTRODUCTION

In the late 1990s, the availability of powerful computing capability, large storage devices, high-speed networking and especially the advent of the Internet, led to a phenomenal growth of digital multimedia content in terms of size, diversity and impact. As suggested by its name, "multimedia" is a name given to a collection of multiple types of data, which include not only "traditional multimedia" such as images and videos, but also emerging media such as 3D graphics (like VRML objects) and Web animations (like Flash animations). Furthermore, multimedia techniques have been penetrating into a growing number of applications, ranging from document-editing software to digital libraries and many Web applications. For example, most people who have used Microsoft Word have tried to insert pictures and diagrams into their documents, and they have the experience of watching online video clips, such as movie trailers. In other words, multimedia data have been in every corner of the digital world. With the huge volume of multimedia data, finding and accessing the multimedia documents that satisfy people's needs in an accurate and efficient manner became a non-trivial problem. This problem is defined as multimedia information retrieval.

The core of multimedia information retrieval is to compute the degree of relevance between users' information needs and multimedia data. A user's information need is expressed as a query, which can be in various forms, such as a line of free text like, "Find me the photos of George

Washington"; a few key words, like, "George Washington photo"; or a media object, like a picture of George Washington. Moreover, the multimedia data are also represented by a certain form of summarization, typically called an index, which is directly matched against queries. Similar to a query, the index can take a variety of forms, including key words and features such as color histograms and motion vectors, depending on the data and task characteristics.

For textual documents, mature information retrieval (IR) technologies have been developed and successfully applied in commercial systems such as Web search engines. In comparison, the research on multimedia retrieval is still in its early stage. Unlike textual data, which can be well represented by key words as an index, multimedia data lack an effective, semantic-level representation (or index) that can be computed automatically, which makes multimedia retrieval a much harder research problem. On the other hand, the diversity and complexity of multimedia offer new opportunities for its retrieval task to be leveraged by the state of the art in various research areas. In fact, research on multimedia retrieval has been initiated and investigated by researchers from areas of multimedia database, computer vision, natural language processing, human-computer interaction and so forth. Overall, it is currently a very active research area that has many interactions with other areas.

In the following sections, we will overview the techniques for multimedia information retrieval and review the applications and challenges in this area. Then, future trends will be discussed. Some important terms in this area are defined at the end of this article.

MULTIMEDIA RETRIEVAL TECHNIQUES

Despite the various techniques proposed in literature, there exist two major approaches to multimedia retrieval; namely, text-based and content-based. Their main difference lies in the type of index: The former approach uses text (key words) as the index, whereas the latter uses low-level features extracted from multimedia data. As a result, they differ from each other in many other aspects, ranging from feature extraction to similarity measurement.

Text-Based Multimedia Retrieval

Text-based multimedia retrieval approaches apply mature IR techniques to the domain of multimedia retrieval. A typical text-IR method matches text queries posed by users with descriptive key words extracted from documents. To use the method for multimedia, textual descriptions (typically key word annotations) of the multimedia objects need to be extracted. Once the textual descriptions are available, multimedia retrieval boils down to a text-IR problem. In early years, such descriptions were usually obtained by manually annotating the multimedia data with key words (Tamura & Yokoya, 1984). Apparently, this approach is not scalable to large datasets, due to its labor-intensive nature and vulnerability to human biases. There also have been proposals from computer vision and pattern recognition areas on automatically annotating the images and videos with key words based on their low-level visual/audio features (Barnard, Duygulu, Freitas, Forsyth, Blei, D. & Jordan, 2003). Most of these approaches involve supervised or unsupervised machine learning, which tries to map low-level features into descriptive key words. However, due to the large gap between the multimedia data form (e.g., pixels, digits) and their semantic meanings, it is unlikely to produce high-quality key word annotations automatically. Some of the systems are semi-automatic, attempting to propagate key words from a set of initially annotated objects to other objects. In other applications, descriptive key words can be easily accessible for multimedia data. For example, for images and videos embedded in

Web pages, the text surrounding them is usually a good description, which has been explored in the work of Smith and Chang (1997).

Since key word annotations can precisely describe the semantic meanings of multimedia data, the text-based retrieval approach is effective in terms of retrieving multimedia data that are *semantically relevant* to the users' needs. Moreover, because many people find it convenient and effective to use text (or key words) to express their information requests, as demonstrated by the fact that most commercial search engines (e.g., Google) support text queries, this approach has the advantage of being amenable to average users. But the bottleneck of this approach is still on the acquisition of key word annotations, since there are no indexing techniques that guarantee both efficiency and accuracy if the annotations are not directly available.

Content-Based Multimedia Retrieval

The idea of content-based retrieval first came from the area of content-based image retrieval (CBIR) (Flickner, Sawhney, Niblack, Ashley, Huang, Dom, Gorkani, Hafner, Lee, Petkovic, Steele & Yanker, 1995; Smeulders, Worring, Santini, Gupta & Jain, 2000). Gradually, the idea has been applied to retrieval tasks for other media types, resulting in content-based video retrieval (Hauptmann et al., 2002; Somliar, 1994) and content-based audio retrieval (Foote, 1999). The word "content" here refers to the bottom-level representation of the data, such as pixels for bitmap images, MPEG bit-streams for MPEG-format video and so forth. Content-based retrieval, as opposed to a text-based one, exploits the features that are (automatically) extracted from the low-level representation of the data, usually denoted as low-level features since they do not directly capture the high-level meanings of the data. (In a sense, text-based retrieval of documents is also "content based," since key words are extracted from the content of documents.) The low-level features used for retrieval depend on the specific data type: A color histogram is a typical feature for image retrieval, motion vector is used for video retrieval, and so forth. Despite the heterogeneity of the features, in most cases, they can be transformed into feature vector(s). Thus, the similarity between media objects can be measured by the distance of their respective feature vectors in the vector space under certain distance metrics. Various distance measures, such as Euclidean distance and M-distance, can be used as the similarity metrics. This has a correspondence to the vector-based model for (text) information retrieval, where a bag of key words is also represented as a vector.

Content-based retrieval also influences the way a query is composed. Since a media object is represented by its low-level feature vector(s), a query must be also transformed into a feature vector to match against the object. This results in query-by-example (QBE) (Flickner et al., 1995), a new search paradigm where media objects such as images or video clips are used as query examples to find other objects similar to them, where "similar" is defined mainly at perceptual levels (i.e., looks like or sounds like). In this case, feature vector(s) extracted from the example object(s) are matched with the feature vectors of the candidate objects. A vast majority of content-based retrieval systems use QBE as its search paradigm. However, there are also content-based systems that use alternative ways to let users specify their intended low-level features, such as by selecting from some templates or a small set of feature options (i.e., "red," "black" or "blue").

The features and similarity metrics used by many content-based retrieval systems are chosen heuristically and are therefore ad-hoc and unjustified. It is very questionable that the features and metrics are optimal or close to optimal. Thus, there have been efforts seeking for theoretically justified retrieval approaches whose optimality is guaranteed under certain circumstances. Many of these approaches treat retrieval as a machine-learning problem of finding the most

effective (weighted) combination of features and similarity metrics to solve a particular query or set of queries. Such learning can be done online in the middle of the retrieval process, based on user-given feedback evaluations or automatically derived "pseudo" feedback. In fact, relevance feedback (Rui, Huang, Ortega & Mehrotra, 1998) has been one of the hot topics in content-based retrieval. Off-line learning has also been used to find effective features/weights based on previous retrieval experiences. However, machine learning is unlikely to be the magic answer for the content-based retrieval problem, because it is impossible to have training data for basically an infinite number of queries, and users are usually unwilling to give feedback.

Overall, content-based retrieval has the advantage of being fully automatic from the feature extraction to similarity computation, and thus scalable to real systems. With the QBE search paradigm, it is also able to capture the perceptual aspects of multimedia data that cannot be easily depicted by text. The downside of content-based retrieval is mainly due to the so-called "semantic gap" between low-level features and the semantic meanings of the data. Given that users prefer semantically relevant results, content-based methods suffer from the low precision/recall problem, which prevents them from being used in commercial systems. Another problem lies in the difficulty of finding a suitable example object to form an effective query if the QBE paradigm is used.

APPLICATIONS AND CHALLENGES

Though far from mature, multimedia retrieval techniques have been widely used in a number of applications. The most visible application is on Web search engines for images, such as the Google Image search engine (Brin & Page, 1998), Ditto.com and so forth. All these systems are text-based, implying that a text query is a better vehicle of users' information need than an example-based query. Content-based retrieval is not applicable here due to its low accuracy problem, which gets even worse due to the huge data volume. Web search engines acquire textual annotations (of images) automatically by analyzing the text in Web pages, but the results for some popular queries may be manually crafted. Because of the huge data volume on the Web, the relevant data to a given query can be enormous. Therefore, the search engines need to deal with the problem of "authoritativeness" – namely, determining how authoritative a piece of data is – besides the problem of relevance. In addition to the Web, there are many digital libraries, such as Microsoft Encarta Encyclopedia, that have the facilities for searching multimedia objects like images and video clips by text. The search is usually realized by matching manual annotations with text queries.

Multimedia retrieval techniques have also been applied to some narrow domains, such as news videos, sports videos and medical imaging. NIST TREC Video Retrieval Evaluation has attracted many research efforts devoted to various retrieval tasks on broadcast news video based on automatic analysis of video content. Sports videos, like basketball programs and baseball programs, have been studied to support intelligent access and summarization (Zhang & Chang, 2002). In the medical imaging area, for example, Liu et al. (2002) applied retrieval techniques to detect a brain tumor from CT/MR images. Content-based techniques have achieved some level of success in these domains, because the data size is relatively small, and domain-specific features can be crafted to capture the idiosyncrasy of the data. Generally speaking, however, there is no killer application where content-based retrieval techniques can achieve a fundamental breakthrough.

The emerging applications of multimedia also raise new challenges for multimedia retrieval technologies. One such challenge comes from the new media formats emerged in recent years, such as Flash animation, PowerPoint file and Synchro-

nized Multimedia Integration Language (SMIL). These new formats demand specific retrieval methods. Moreover, their intrinsic complexity (some of them can recursively contain media components) brings up new research problems not addressed by current techniques. There already have been recent efforts devoted to these new media, such as Flash animation retrieval (Yang, Li, Liu & Zhuang, 2002a) and PowerPoint presentation retrieval. Another challenge rises from the idea of retrieving multiple types of media data in a uniform framework, which will be discussed next.

FUTURE TRENDS

In a sense, most existing multimedia retrieval methods are not genuinely for "multimedia," but are for a specific type (or modality) of non-textual data. There is, however, the need to design a real "multimedia" retrieval system that can handle multiple data modalities in a cooperative framework. First, in multimedia databases like the Web, different types of media objects coexist as an organic whole to convey the intended information. Naturally, users would be interested in seeing the complete information by accessing all the relevant media objects regardless of their modality, preferably from a single query. For example, a user interested in a new car model would like to see pictures of the car and meanwhile read articles on it. Sometimes, depending on the physical conditions, such as networks and displaying devices, users may want to see a particular presentation of the information in appropriate modality(-ies). Furthermore, some data types, such as video, intrinsically consist of data of multiple modalities (audio, closed-caption, video images). It is advantageous to explore all these modalities and let them complement each other to obtain a better retrieval effect. To sum, a retrieval system that goes across different media types and integrates multi-modality information is highly desirable.

Informedia (Hauptmann et al., 2002) is a well-known video retrieval system that successfully combines multi-modal features. Its retrieval function not only relies on the transcript generated from a speech recognizer and/or detected from overlaid text on screen, but also utilizes features such as face detection and recognition results, image similarity and so forth. Statistical learning methods are widely used in Informedia to intelligently combine the various types of information. Many other systems integrate features from at least two modalities for retrieval purpose. For example, the WebSEEK system (Smith & Chang, 1997) extracts key words from the surrounding text of images and videos in Web pages, which is used as their indexes in the retrieval process. Although the systems involve more than one media type, typically, textual information plays the vital role in providing the (semantic) annotation of the other media types.

Systems featuring a higher degree of integration of multiple modalities are emerging. More recently, MediaNet (Benitez, Smith & Chang, 2002) and multimedia thesaurus (MMT) (Tansley, 1998) are proposed, both of which seek to provide a multimedia representation of a semantic concept – a concept described by various media objects including text, image, video and so forth – and establish the relationships among these concepts. MediaNet extends the notion of relationships to include even perceptual relationships among media objects.

Yang, Li and Zhuang (2002b) propose a very comprehensive and flexible model named *Octopus* to perform an "aggressive" search of multi-modality data. It is based on a multi-faceted knowledge base represented by a layered graph model, which captures the relevance between media objects of any type from various perspectives, such as the similarity on low-level features, structural relationships such as hyperlinks and semantic relevance. Link analysis techniques can be used to find the most relevant objects for any given object in the graph. This new model can accommodate

knowledge from various sources, and it allows a query to be composed flexibly using either text or example objects, or both.

CONCLUSION

Multimedia information retrieval is a relatively new area that has been receiving more attention from various research areas like database, computer vision, natural language and machine learning, as well as from industry. Given the continuing growth of multimedia data, research in this area will expectedly become more active, since it is critical to the success of various multimedia applications. However, technological breakthroughs and killer applications in this area are yet to come, and before that, multimedia retrieval techniques can hardly be migrated to commercial applications. The breakthrough in this area depends on the joint efforts from its related areas, and therefore, it offers researchers opportunities to tackle the problem from different paths and with different methodologies.

REFERENCES

Barnard, K., Duygulu, P., Freitas, N., Forsyth, D., Blei, D., & Jordan, M. (2003). Matching words and pictures. *Journal of Machine Learning Research, 3,* 1107-1135.

Benitez, A.B., Smith, J.R., & Chang, S.F. (2000). MediaNet: A Multimedia Information Network for knowledge representation. *Proceedings of the SPIE 2000 Conference on Internet Multimedia Management Systems,* 4210.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Proceedings of the 7th International World Wide Web Conference,* 107-117.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by image and video content: The QBIC system. *IEEE Computer, 28*(9), 23-32.

Foote, J. (1999). An overview of audio information retrieval. Multimedia Systems, 7(1), 2-10.

Hauptmann, A., et al. (2002). Video classification and retrieval with the Informedia Digital Video Library System. *Text Retrieval Conference (TREC02),* Gaithersburg, MD.

Liu, Y., Lazar, N., & Rothfus, W. (2002). Semantic-based biomedical image indexing and retrieval. *International Conference on Diagnostic Imaging and Analysis (ICDIA 2002).*

Lu, Y, Hu, C., Zhu, X., Zhang, H., Yang Q., (2000). A unified framework for semantics and feature based relevance feedback in image retrieval systems. *Proceedings of ACM Multimedia Conference,* 31-38.

NIST TREC Video Retrieval Evaluation. Retrieved from *www-nlpir.nist.gov/projects/trecvid/*

Rui, Y., Huang, T.S., Ortega, M., & Mehrotra, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval. IEEE Trans on Circuits and Systems for Video Technology *(Special Issue on Segmentation, Description, and Retrieval of Video Content) 8,* 644-655.

Smeulders, M., Worring, S., Santini, A., Gupta, & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349-1380.

Smith, J.R., & Chang, S.F. (1997). Visually searching the Web for content. *IEEE Multimedia Magazine, 4*(3), 12-20.

Somliar, S.W., Zhang, H., et al. (1994). Content-based video indexing and retrieval. *IEEE MultiMedia, 1*(2), 62-72.

Synchronized Multimedia Integration Language (SMIL). Retrieved from *www.w3.org/AudioVideo/*

Tamura, H., & Yokoya, N. (1984) Image database systems: A survey. *Pattern Recognition*, *17*(1), 29-43.

Tansley, R. (1998). The Multimedia Thesaurus: An aid for multimedia information retrieval and navigation (masters thesis). *Computer Science*, University of Southampton.

Yang, J., Li, Q., Liu W., & Zhuang, Y. (2002a). FLAME: A generic framework for content-based Flash retrieval. *ACM MM'2002 Workshop on Multimedia Information Retrieval*, Juan-les-Pins, France.

Yang, J., Li, Q., & Zhuang, Y. (2002b). Octopus: Aggressive search of multi-modality data using multifaceted knowledge base. *Proceedings of the 11th International Conference on World Wide Web*, 54-64.

Zhang, D., & Chang, S.F (2002). Event detection in baseball video using superimposed caption recognition. *Proceedings of ACM Multimedia Conference*, 315-318.

KEY TERMS

Content-Based Retrieval: An important retrieval method for multimedia data, which uses the low-level features (automatically) extracted from the data as the indexes to match with queries. Content-based image retrieval is a good example. The specific low-level features used depend on the data type: Color, shape and texture features are common features for images, while kinetic energy and motion vectors are used to describe video data. Correspondingly, a query also can be represented in terms of features so that it can be matched against the data.

Index: In the area of information retrieval, an "index" is the representation or summarization of a data item used for matching with queries to obtain the similarity between the data and the query, or matching with the indexes of other data items. For example, key words are frequently used indexes of textual documents, and color histogram is a common index of images. Indexes can be manually assigned or automatically extracted. The text description of an image is usually manually given, but its color histogram can be computed by programs.

Information Retrieval (IR): The research area that deals with the storage, indexing, organization of, search, and access to information items, typically textual documents. Although its definition includes multimedia retrieval (since information items can be multimedia), the conventional IR refers to the work on textual documents, including retrieval, classification, clustering, filtering, visualization, summarization and so forth. The research on IR started nearly half a century ago and it grew fast in the past 20 years with the efforts of librarians, information experts, researchers on artificial intelligence and other areas. A system for the retrieval of textual data is an IR system, such as all the commercial Web search engines.

Multimedia Database: A database system dedicated to the storage, management and access of one or more media types, such as text, image, video, sound, diagram and so forth. For example, an image database such as Corel Image Gallery that stores a large number of pictures and allows users to browse them or search them by key words can be regarded as a multimedia database. An electronic encyclopedia such as Microsoft Encarta Encyclopedia, which consists of tens of thousands of multimedia documents with text descriptions, photos, video clips and animations, is another typical example of a multimedia database.

Multimedia Document: A multimedia document is a natural extension of a conventional textual document in the multimedia area. It is defined as a digital document composed of one or multiple media elements of different types (text, image, video, etc.) as a logically coherent unit. A multimedia document can be a single picture or a single MPEG video file, but more often it is a complicated document, such as a Web page, consisting of both text and images.

Multimedia Information Retrieval (System): Storage, indexing, search and delivery of multimedia data such as images, videos, sounds, 3D graphics or their combination. By definition, it includes works on, for example, extracting descriptive features from images, reducing high-dimensional indexes into low-dimensional ones, defining new similarity metrics, efficient delivery of the retrieved data and so forth. Systems that provide all or part of the above functionalities are multimedia retrieval systems. The Google image search engine is a typical example of such a system. A video-on-demand site that allows people to search movies by their titles is another example.

Multi-Modality: Multiple types of media data, or multiple aspects of a data item. Its emphasis is on the existence of more than one type (aspects) of data. For example, a clip of digital broadcast news video has multiple modalities, include the audio, video frames, closed-caption (text) and so forth.

Query-by-Example (QBE): A method of forming queries that contains one or more media object(s) as examples with the intention of finding similar objects. A typical example of QBE is the function of "See Similar Pages" provided in the Google search engine, which supports finding Web pages similar to a given page. Using an image to search for visually similar images is another good example.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 710-716, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.21
Multimedia Capture, Collaboration and Knowledge Management

Subramanyam Vdaygiri
Siemens Corporate Research Inc., USA

Stuart Goose
Siemens Corporate Research Inc., USA

ABSTRACT

This chapter presents methods and technologies from Siemens Corporate Research that can assist in the process of creating multimedia collaborative knowledge bases: capture, querying, visualization, archiving, and reusability of multimedia knowledge bases. A selection of Siemens products in the healthcare and communication domains are introduced, above which novel multimedia collaboration and knowledge management technologies have been developed by the authors. With examples, it is explained how in concert these technologies can contribute to streamlining the processes within healthcare enterprises, telemedicine environments and home healthcare practices.

INTRODUCTION AND MOTIVATION

The networked healthcare enterprise is providing unprecedented opportunities for healthcare workers to collaborate and make clinical decisions in an efficient manner. Significant progress has been made to enable healthcare personnel to obtain answers to simple clinical questions by using medical databases of evidence-based answers. This approach allows reuse and sharing of knowledge to help healthcare professionals to save time and effort and help patients in an efficient manner. For situations where healthcare workers have simple questions with simple answers, this approach is perhaps overkill. However, when the questions are of a more complex nature, by capturing and archiving complex answers in a rich multimedia form they can be exploited multiple in the future

to explain solutions in a manner that can be easily digested by healthcare workers.

A contemporary healthcare enterprise involves complex media elements (images, videos, documents, etc.) and volumes of documentation both digital and on paper. The healthcare knowledge base should incorporate these media elements and easily allow users to search, extract, and reuse. Some of the modern knowledge management systems allow building of communities of practice around documents. But there is a need to move beyond regular office documents to address rich media and encompass specialized medical and clinical data.

The networked enterprise is also enabling a plethora of ways for healthcare personnel to communicate and collaborate. The next generation of communication technologies will bring converged voice and data solutions on a single network. This is helping integrate the healthcare IT (Information Technology) systems with Web-based communications. In recent years we have witnessed a proliferation of communication and data devices like GPRS cellular phones and PDAs, thus providing an opportunity for accessing clinical information anywhere/anytime and allowing users to collaborate over clinical information to reach decisions quickly.

The concept of presence and availability offered by various instant messaging tools is changing the manner in which people are communicating with each other. *Presence* enables a user to know who in their contact or buddy list is available or not at any given point in time. *Availability* options allow a user to signal whether they are available to be contacted and which form of communication they favor. Presence and availability information allow users to interact in various ways in offline, real-time or in near-real time modes. Mobile communication technologies are being developed that enables mobile location and presence. The integration of the healthcare enterprise content repository with a Web-based infrastructure and presence and availability represent the three pillars of modern unified, or converged, communication.

Although the potential for a rich communications and IT infrastructure is high, there remains a need to streamline the communications and collaborations between healthcare personnel to ensure that valuable knowledge gained from daily interactions between healthcare personnel is not lost. This chapter presents methods and technologies from Siemens Corporate Research that can assist in the process of creating multimedia collaborative knowledge bases: capture, querying, visualization, archiving, and reusability of multimedia knowledge bases. Throughout the chapter, a number of Web-based technologies are introduced that enable healthcare personnel to interact in a variety of modes regardless of whether they are mobile or sedentary.

BACKGROUND: STREAMLINING HEALTHCARE AND TELEHEALTH

Since the advent of Web-based workflows, there has been a growing emphasis in healthcare enterprises on methods to increase organizational efficiencies, reduce errors and focus on patient care. One such platform is Soarian (Siemens Soarian) from Siemens Medical (Siemens Medical) that offers an integrated workflow technology that can streamline the operational processes of healthcare. Soarian's infrastructure has been engineered based on clinical processes that enable physicians to focus less on administrative duties and more care by providing them with access to all clinical data in a single view. The goals being to offer actionable guidelines to clinicians based on best practices and to support them in reaching accurate, evidence-based decisions promptly.

A nascent area for which technology can be a significant enabler is that of telemedicine (Hibbert, Mair, May, Boland, O'Connor, Capewell, & Angus, 2004), where clinical needs are extended beyond the boundary of the hospital. A key re-

quirement to facilitate a telehealth consultation is to have medical personnel in remote locations to communicate and collaborate with each other in a quick, efficient, and effective manner.

Some of the issues that an integrated communications and healthcare medical IT infrastructure can help address are:

- Reducing time and effort wasted in daily communications (paging, phone calls) between clinicians, nurses, and patients
- Reducing the communication pathways by a combined infrastructure can streamline how users can be reached at a given point in time
- Complex solutions to questions can be assembled in multimedia fashion to succinctly convey the message
- Knowledge exchanged and gained from daily interactions between people can captured and archived with high granularity
- Reuse, by querying and retrieving relevant segments from past collaborative activities, can avoid recreating the wheel when a recurrent problem reappears
- Reduce paper work by having one consolidated IT-communications infrastructure
- Reduce cost by managing a single converged voice and data network

Remote communications in contemporary telehealth systems allows users to collaborate in two different modes: store & forward and in real time. The asynchronous mode, or store & forward method, allows for the sharing of medical information in an offline mode, but the absence of human interactivity prevents healthcare personnel from augmenting this with personal comments, insight and knowledge. The real time collaboration mode requires the session to be scheduled in advance, and all participants must be dedicated to the collaboration session for the duration.

The current mode of document/data sharing involves all users within telehealth WAN (Wide Area Network) cluster to upload documents, data and code to a central repository from where other users can extract materials of interest. This mode of collaboration among users is not sufficient for many users to exchange complex ideas and viewpoints regarding various images and other documents in intricate detail. For example, medical researchers frequently need to locate and reference information that is not only physically distributed across the sites of their collaborators, but also in an array of formats (images, reports generated based on experiments, or code execution).

Often users collaborate by exchanging e-mail messages along with data/documents downloaded from a central Web server. Although this allows individuals to participate at their own pace, users invest much time typing descriptive text to discuss a particular topic, especially when expressing complex thoughts within context of a document or image. The problem becomes particularly acute when users are collaborating using document attachments. This mode leads to large documents being exchanged as attachments back and forth between users, with the potential for inconsistencies to arise between successive versions. Moreover, with clinicians collaborating across time zones on documents or images, a need has been observed for conveying complex information via rich multimedia.

To anticipate and streamline the future communication and IT needs of healthcare enterprises and burgeoning telehealth market, the various requirements were distilled:

- **Communication**: It is clear that there exists a need to combine communications within the healthcare processes to optimize the clinicians' time to manage better the inexorable stream of phone calls, paging, exchange of paper and digital documents. Presence-based communications enables the healthcare personnel to communicate in a very effective manner.

- **Collaboration:** The tools should provide a consistent manner to collaborate irrespective of the mode, all online in a real time conference, or a single user making her analysis and comments on an image in offline mode. By allowing users to collaborate in different modes within a browser-based environment on diverse medical data can facilitate a seamless collaboration workflow. These technologies can enhance typical telemedicine scenarios and also extend conventional doctor-patient interactions to include web-based interactions such as chronic care or follow-ups.
- **Knowledge Management**: There should also be a consistent way to archive the multimedia annotations made on such documents. This would not only allow researchers to collaborate based on their time and schedule, but also to search, retrieve and filter all comments and analysis made previously by collaborators on any multimedia document.

The remaining sections describe ongoing efforts at Siemens Corporate Research to develop technological solutions driven by these anticipated future needs.

TECHNOLOGICAL BUILDING BLOCKS

OpenScape (Siemens OpenScape) from Siemens Communications Networks (Siemens Information and Communications Networks) is a suite of communication applications designed to increase the productivity of information workers. It aims to control the panoply of communication applications and devices, on both fixed and wireless networks, connected via local and wide area networks. OpenScape addresses the fragmented nature of communication modalities and their separation and provides a unifying framework for integrating, managing and streamlining communication in the enterprise.

It obviates redundant communication sessions, where a person calls multiple phone numbers and/or leaves duplicate voice, email, and instant messages in an effort to communicate urgent issues. Hence, it is possible to avoid:

- Unnecessary cell phone intrusions into client meetings, work sessions or personal time;
- Wasted time setting up conference calls, communicating call-in information, sending and synchronizing documents, and establishing separate sessions for voice, Web and video collaboration; and
- The difficulty of mobilizing all key colleagues that may be equipped with different applications, or because setting up a collaborative session is too complex and time consuming.

It provides healthcare personnel the experience of a single, synchronized set of communication resources, sharing common controls and shared communication rules and intelligence. These capabilities can be accessed via a wide range of devices and interfaces to serve the constantly changing needs of mobile employees. OpenScape provides personal and workgroup communications portals for multiple healthcare domains. OpenScape makes extensive use of presence-based communication. There are different ways of showing presence and the user's status:

Device Presence: indicates the presence of an application or communications device. The user could be online or off-line. For example:

- Off hook/on hook status of an IP desk phone or mobile phone
- Instant Messaging (IM) application presence
- Collaboration application presence indicating whether an individual has signed on to the application or not
- On which terminal can the user be reached

User Presence: associates presence with an individual rather than the user's device

- The willingness of the user to be reached
- The activity that the user is currently engaged
- The location of the user—working remotely, out of office, and so on.
- Mood—happy, good or bad tempered or even annoyed

Device presence or terminal status will become less important as terminals become more advanced and intelligent, with the ability to handle a multitude of multimedia content adaptively. A mobile phone or PDA, for instance will be able to negotiate with the remote communication party the type of information it can support, for instance video, audio, and so on.

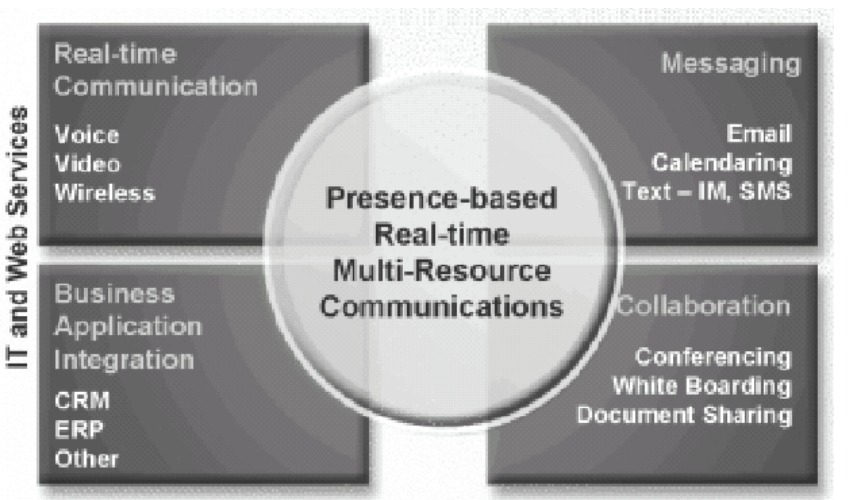
IMS (Siemens IMS) from Siemens Communications Mobile (http://www.siemens-mobile.com) provides new presence and communications platform for mobile devices and networks. The IMS platform was standardized for new multimedia applications and services that could be rapidly deployed by mobile network operators, such as audio/video conferencing, chat, and presence services over new mobile devices. Moreover, IMS is positioned to voice and multimedia communication. Different from the legacy circuit-switched voice/data communication in regular phone calls, IMS is based on IP technology that can control real-time and non-real-time services on the same IP network.

Together OpenScape, IMS and Soarian provide the core technological building blocks that enable the integration of the healthcare enterprise content repository with a Web-based infrastructure and presence and availability. Our research work is layered above these three pillars and the following sections illustrate the benefits of a converged communication healthcare enterprise.

ENRICHING COLLABORATIVE CONSULTATIONS

During a telehealth consultation between patient and clinician various medical documents might be used including laboratory reports, photos of injury/skin lesions, x-rays, pathology slides, EKGs, MR/CT images, medical claim forms, prescriptions, clinical results, case reports, and other documents. Some of the documents like photos and images

Figure 1. OpenScape: A way of communication



might have been captured using camera during patient visits, while other clinical (lab reports) or financial information (insurance claim) were gleaned from other information systems. Face-to-face video conferencing is commonly used in telemedicine not only for personalized remote communications, but also for regulatory reasons as evidence of a consultation the costs for which can subsequently be reimbursed.

It would be convenient if a clinician could combine, interface, convert, and extract disparate medical information such as those listed above from peripheral devices like photo cameras, video conferencing session, content management systems, PACS, and regular office documents into a Web-based *composite* document. Thereafter, it would be convenient if this composite document could be used as the basis for browser-based collaboration (whether it is offline or real time) between various participants. Such a composite document generated should combine all relevant information needed for a particular collaboration session into one seamless document so that effective offline or real time collaboration can occur.

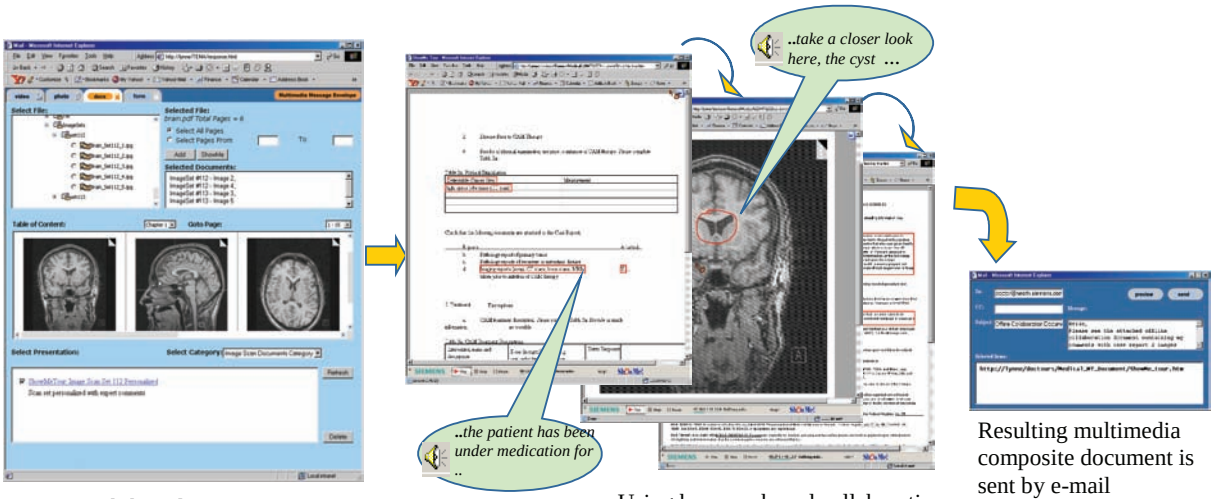
Let us now look at some of the details involved in realizing our solution. A user could choose specific pages from document(s) stored locally and automatically have the selected pages converted to HTML format and hyperlinked with each other to form a composite document. Then, he or she could highlight important parts of the document and add personal comments with the help of voice and graphic annotations using our multimedia presentation software, called *ShowMe* (Sastry, Lewis, & Pizano, 1999). The multimedia annotation technology developed is unique as it not only captures the spatial nature but also the temporal aspects. For instance, the multimedia annotation on a document would capture a synchronized temporal voice, graphic and mouse pointer annotations. Finally, the user could save the composite document along with the annotations on the local web server, and this document is referred to as the collaboration document. Associated with this collaboration document is some metadata in the form of an XML schema that describes this document. This metadata is finally uploaded to the central server and its URL on this central server can then be sent to other participants.

Participants can view the collaboration document via the URL of the metadata stored on the central Web server using a lightweight Web-browser player. As the collaboration documents are stored locally, they are amenable to document management tasks, including deleting, moving, and so on. However, any such document management task must be accompanied by making an appropriate change in the metadata located at the central Web server. The above process allows various users to collaborate over documents quickly and easily by only sharing information relevant to the topic in question. In addition, there is increase in productivity as users can quickly exchange information without having to exchange e-mail to explain problem/solution.

Figures 2, 3, and 4 show a particular workflow implemented to demonstrate the spectrum of modes of browser-based collaboration using office documents and Web content. The components can be re-used in several other collaboration workflows and processes. As only specific parts of different documents might be needed for a particular collaborative session, participants are able to combine, on the fly, specific pages from several documents with different formats into one seamless Web-based composite document. Using synchronous voice and graphic annotations along with mouse as a pointer, a clinician can continuously narrate a patient's case history across the entire composite document while using the mouse to gesture or annotate on different parts of the aggregated document.

These modes of collaboration—whether offline, IM-based, or real-time—take place using a regular web browser with the requisite collaboration plug-ins. This allows the participants to collaborate within the familiar environment

Figure 2. Multimedia Enhanced Store & Forward Collaboration: A clinician can quickly create one seamless web-based composite multimedia document by combining various medical information segments like images, photos, EKGs together. Synchronous voice, graphic and mouse annotations can then be easily added on top of the composite document before sending it via regular e-mail. The recipient needs only a web browser to view the composite document along with the voice and graphic annotations.



Web-based Composite Document Creation

Using browser based collaboration tools, the user narrates and annotates

Resulting multimedia composite document is sent by e-mail

Figure 3. Instant Messaging Based Document Collaboration: This uses presence and availability of different participants to setup collaboration. Also, this allows asynchronous messaging of voice and graphic annotations within a real-time collaboration session. This enables users to exchange information at one's own pace and yet participate in a near real-time collaboration session.

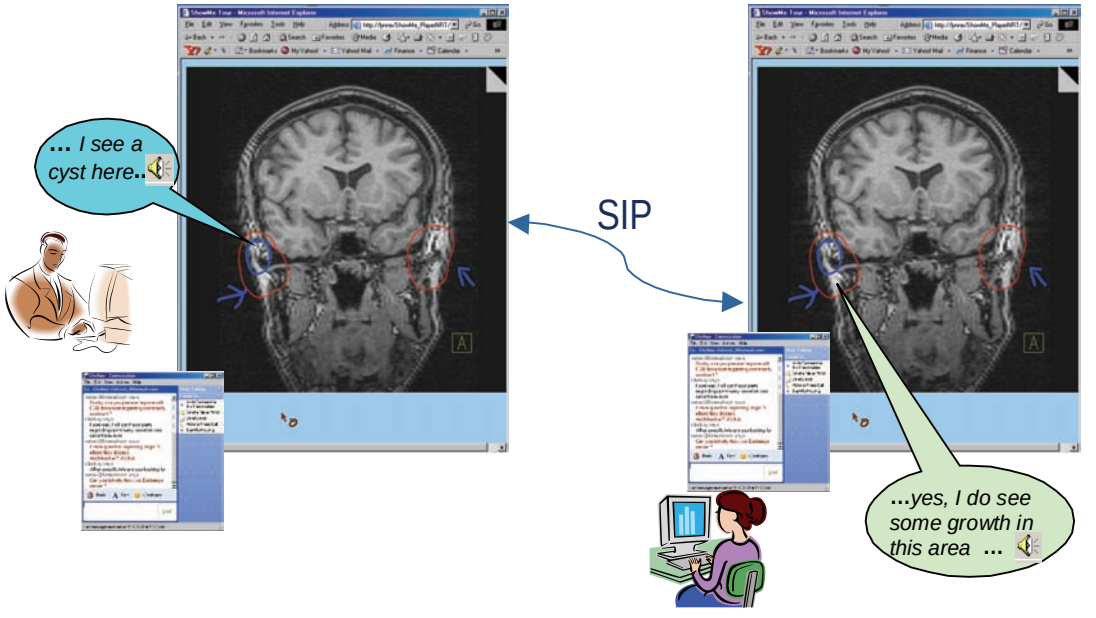
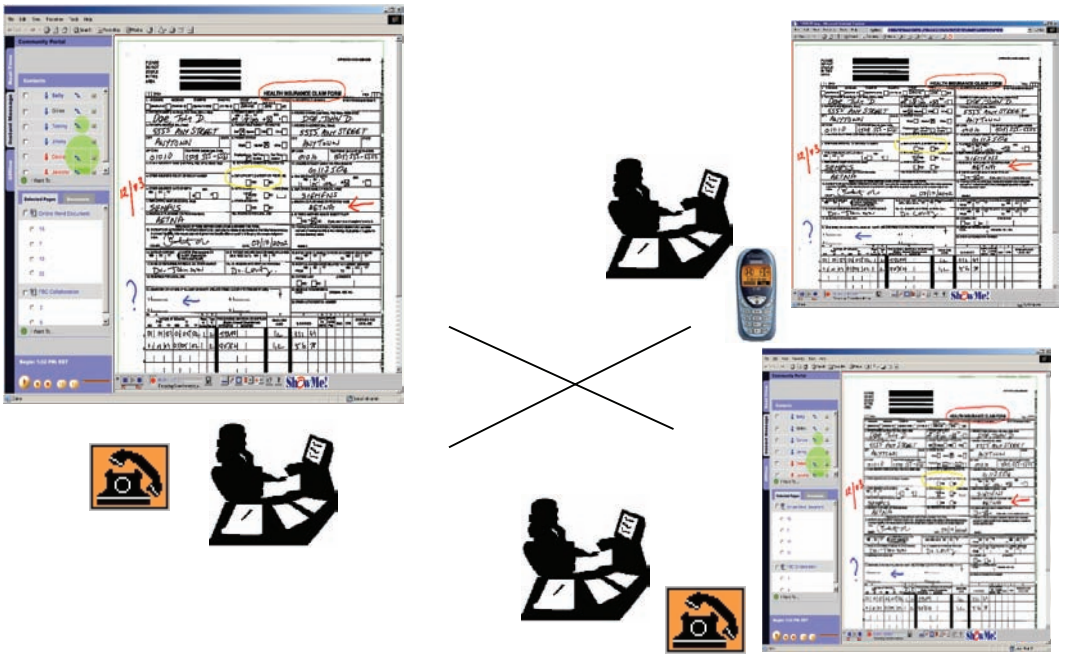


Figure 4. Multi-participant Real-Time Collaboration: Several remote participants using their web browser can collaborate on the generated composite medical document. The collaboration is accompanied by a voice conference.



of a Web browser. The annotations by various participants during a collaboration session on composite medical documents can be archived into a database in a very lightweight manner. The voice and graphic annotations along with the composite document can be archived with high granularity along with meta-data describing the various annotations. This would allow for searching and retrieval of not only medical information/documents exchanged during collaboration sessions, but also to obtain information regarding comments, interactions and dialogue between various participants. For instance, one can query and retrieve specific comments/annotations made by particular participant from a past collaboration session on a specific medical document. In addition, one can easily follow the changes in the medical condition of the patient by comparing the medical documents and associated annotations made during successive patient visits.

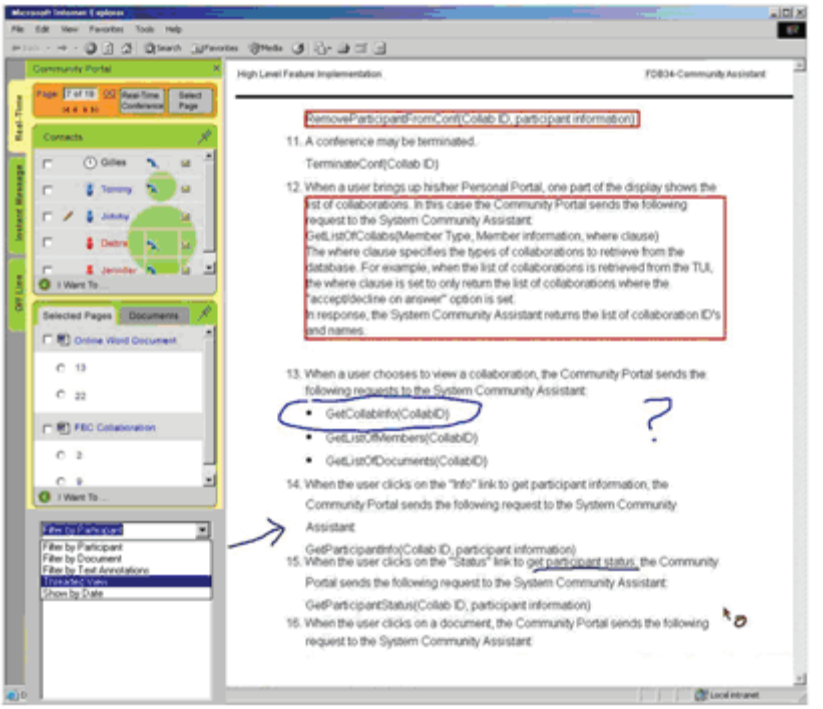
COLLABORATION KNOWLEDGE ACCESS AND MANAGEMENT

Access to and the management of the knowledge contained within the collaboration documents introduced above is reviewed in this section.

Filtering Collaboration Annotations

The ability to save annotations from a real time, offline or IM based document collaboration enables participants to filter and view a conference in several different ways. The figure below illustrates how an archived conference may be viewed in different ways by applying different filters. For example, the filters can be to show the conference view by participants, by documents collaborated on, by threaded view, by type of annotations, by session, and so on. New filters can be easily created based on the use case scenario.

Figure 5. The clinician can filter and quickly access specific segments from past collaborative sessions and reuse to address the current problem.



Navigation, Access and Reusability of Knowledge Documents

The ability to archive annotations (both temporal and spatial) allows the users to retrieve and reuse segments of existing annotated documents from any mode of collaboration. For example, during a real time conference, several participants collaborated on different composite pages (where the composite pages might have come from different original documents). A user can select a particular page from an existing collaboration document, filter annotations based on any criteria (as described above), add any more composite pages, add more temporal/spatial annotations and collaborate in off-line, IM, or real time mode. This allows the enterprise to preserve the knowledge gathered during several collaboration sessions performed in any mode, while still enabling users to select parts of annotated documents from previous sessions for further reuse in future collaboration sessions.

Multimedia Response and Discussion Boards

The structure of the collaboration archive allows one user to respond to any particular comment or annotation made by another user. The response could be in the form of a temporal or spatial annotation, or both. One scenario is that a sender can compose a composite document, make his or her annotations and send to the recipient(s). The annotated document could be sent as an attachment, or accessed from a central server. In the latter case, the annotated document is sent as a URL link. The recipient(s) can overlay their replies (annotations) and send back the annotated document, or the annotated document can be further annotated by other users (forwarding via e-mail.) By quickly overlaying temporal or spatial annotations it allows users to imbue information and temporal context to their responses.

Figure 6. A multimedia threaded discussion board

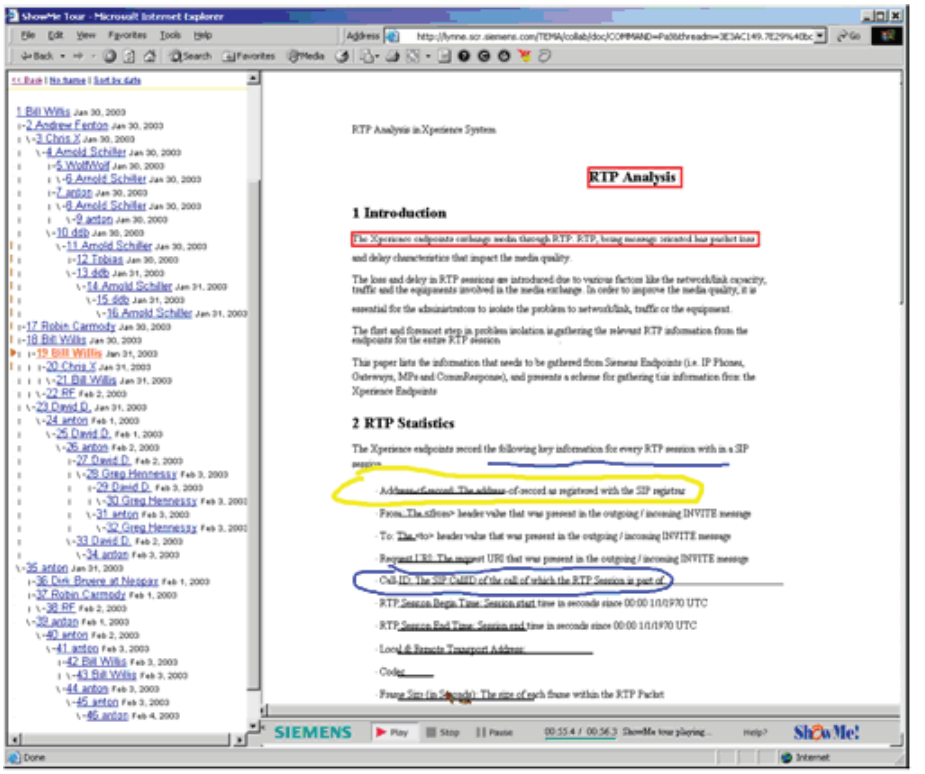


Figure 6 shows document(s) being collaborated upon by several users in different collaboration modes using a threaded approach. These annotations could have been made in offline mode by different users or some of the annotations could have happened in a real time or IM conference and then imported into the discussion thread. For example, a user creates a composite document, adds various annotations in offline mode and sends it to her workgroup. The workgroup meets later and discusses in real time conference mode where annotations are captured as temporally with high granularity. During the real-time conference, they add more composite pages and perform collaborations above them. When finished, one user adds the resulting composite document with the temporal and spatial annotations to the discussion group. Subsequently, participants can add/edit/delete more comments (or annotations), or respond to any particular comment—hence a rich threaded discussion is facilitated. This scenario illustrates how group of users can work in a newsgroup discussion manner and leverage rich multimedia to express their thoughts.

Query and Retrieval of Multimedia Knowledge Base

As reported in the previous sections, the collaboration archive of interactions—voice, graphic, text, or mouse pointer annotations—between participants over several documents in different collaboration modes being captured and archived in a structured manner over time. As knowledge in the form of annotations is captured and reused in high granularity, there is a need for a simple search mechanism to locate appropriate segments of annotated documents from a large archive of annotated composite documents.

Queries to the knowledge base can be based on various criteria, for example, using metadata captured during collaboration session, the annotations using voice, text and the document content, and also by using feature relative annotations one can get the textual content of the document on which annotations were overlaid.

Representation, Visualization and Browsing of Collaborations

To leverage the knowledge contained within archived collaboration sessions, we have also explored various approaches for representing, visualizing and browsing archived collaboration sessions. As collaborations often involve two dimensional documents and are by nature temporal, we are investigating representing the structure and content of collaborations in a three dimensional space. Arguably the most prevalent standard in the 3D graphics arena is VRML: a high-level textual language for describing the geometry and behavior of 3D scenes. Fortunately, a plethora of browser plug-ins are readily available. Following an iterative design/evaluation cycle, we converged on the design of a single VRML model with an interface that embodied and combined many of the characteristics that we sought:

- A viewpoint providing a visual overview/ summary of the session
- A viewpoint providing a detail view of the session
- Interactivity and browsing capabilities at both overview and detail levels

Figures 7 and 8 show two viewpoints on the same VRML model, with each viewpoint having a distinct purpose.

The viewpoint shown in Figure 8 conveys the structure of the session (three topics were discussed), the documents used in the course of the discussion (indicated by the thumbnail icons on the right of the timeline), and that two action items were assigned (the red columns on the right of the timeline).

The viewpoint shown in Figure 8 exposes more detail regarding the content of the session. An audio waveform running along the timeline illustrates the presence of noise and is also color-indexed to the participants. An exploded view of the current document under discussion is show in the middle of the visualization. A small portrait image of the current speaker is shown in the corner of the current document, along with an icon that can be clicked on to initiate a phone call, e-mail, or instant message to that person. In addition, red columns along the left side of the timeline indicate assigned action items with a portrait image of the person responsible.

A VCR metaphor is provided to assist with navigation, and a VCR control panel can be seen at the right of the interface. When the play button is activated, the exploded view of the current

Figure 7. A visual overview of a session that conveys structure, documents used, and action items assigned

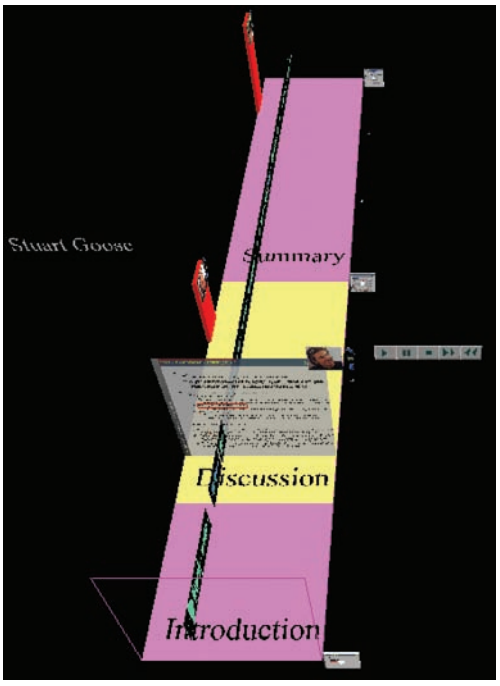
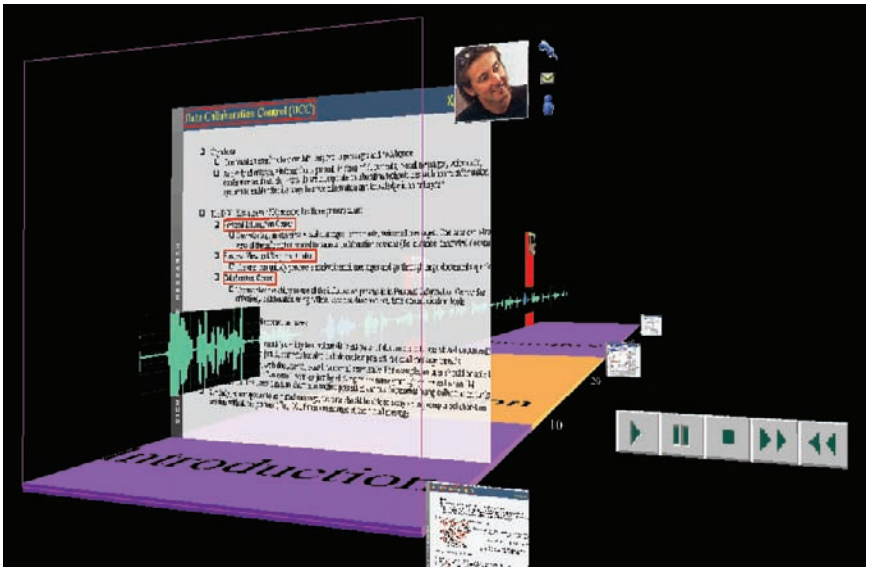


Figure 8. The detail view of a session



document begins to glide backwards through time in synchrony with the audio stream played at normal speed. The forward and rewind buttons move the current position to the beginning of the respective session topic.

IN SUPPORT OF HOME HEALTH: NOMADIC NURSE

Home healthcare nursing is an increasingly widespread type of healthcare that typically involves a nurse driving to a patient's place of residence to provide the necessary monitoring and treatment. There are a wide variety of cases in which home healthcare is recommended by physicians, and recent years have seen a steady increase in the need for this type of care (Giles, 1996). While many industry observers believe home healthcare will increase in the future, many current home healthcare providers are either non-profit agencies or operate with conservative profit margins. In addition, as nurses' time is a scarce resource and IT spending limited, any new technologies that are not economically priced or that require nurses to increase the time spent with patients have a high probability of being passed over or abandoned.

Home healthcare nursing is by nature a peripatetic profession, and as such nurses need to transport their equipment from one home to the next. From our studies described previously, it was observed that the preponderance of nurses used paper and pencil notes augmented by their memory to record the interactions with their patients throughout their working day. The written and memorized notes were entered into the computer-based system upon their return home or to the office. There were a plethora of reasons as to why some nurses chose to operate in this mode. For example, some asserted that anything that breaks the contact between the clinician and the patient breaks the treatment involved in the visit. "There is a healing involved in the physical touch between the nurse and the patient" was the expression used. To avoid breaking this important connection, the application would need to use less obtrusive techniques that still provide for sterile-hands operation. In addition, the application would need to integrate well with the nurses'

workflow and their frequent need for non-linear information access.

Hardware and Software Selection

The hardware selection was influenced by the following factors:

- Laptops were identified as the device of choice for the majority of home healthcare clinicians interviewed during the study. Lowering costs of laptops make them relatively affordable even for low-budget agencies; lightweight of the devices is appropriate in context of patient visits.
- The study revealed that clinicians were open to the idea of using a PDA for reasons of size, weight, and no boot-up latency. However, concern was expressed as to whether the screen size would be adequate for their needs.
- New government regulations (HIPPA) place significant emphasis on ensuring secure access to sensitive patient information in order to preserve patient privacy. Towards satisfying the HIPPA regulations, a Siemens biometric mouse was incorporated (Siemens Biometric Mouse).
- To reduce the impact of the interaction upon the nursing tasks, a discreet wireless Bluetooth earpiece was incorporated. This enables the nurse to issue spoken commands, but also to receive spoken feedback without the patient hearing. The absence of wires was crucial so as not to inhibit the nursing duties.

The software technology selection was influenced by the following factors:

- Clinicians may access the application from a variety of different locations, including the office, their homes, patient homes, from the car while driving, etc.
- Administering clinical care requires sterile conditions, which makes traditional input devices unsuitable. This requirement indicates a multimodal interface (Multimodal Interaction Working Group), which, if necessary, can be controlled entirely by voice.
- Home healthcare practitioners would prefer continuous access to the patient database, however the existing wireless coverage of rural areas remains too fragmented to be reliable. The application needs to be able to work in either offline or online modes of connectivity.

Collectively, these requirements led us to conclude that a WWW-based solution was feasible, and that recent advances in multimodal technologies provide support on various devices. As such, the user interface was developed using a combination of HTML, Java and Javascript.

Notebook Design: Capturing Patient Vital Signs

The goal of our prototype is to show how cost-effective technology can be integrated into the workflow of a nurse to prove that patient vital signs can be captured unobtrusively. If successful, this could obviate the need for the nurse to enter this data into a computer system upon returning to the office.

In order to address the majority of agencies that expressed interest in notebook and PDA platforms, our design sought to offer a similar user experience while attempting to leverage the respective advantages of each. To provide hands-free operation, the nurse's notebook computer is equipped with a Bluetooth (Bluetooth) capability. Bluetooth wireless headset (Bluetooth headset from Siemens) supports mobile speech interaction with the application within an adequate radius. The initial implementation was developed for a notebook, as seen in Figure 9.

As can be seen from Figures 10 and 11, a multimodal interface allows the nurse to use the keyboard/mouse and/or speech to navigate and enter values into the visualization. For the laptop, SALT (SALT Forum) was selected as the technology used to develop the multimodal interface. In the classical SALT paradigm, speech recognition is initiated using either a keyboard or a mouse, but to offer truly hands-free operations we introduced some novel approaches to support continuous recognition and pause/resume functionality. This enables nurses to use verbal commands to indicate to the application that they are about to start dictating commands, or whether they are engaged in the conversation with a patient. Additionally, after temporarily deactivating the speech recognition, the nurse can resume working by simply issuing a voice command.

Figure 9. Nurse's laptop, Bluetooth PCMCIA card and Bluetooth enabled headset

PDA Design: Less is More

The notebook implementation was heavily leveraged for the subsequent implementation for the Pocket PC PDA. While the functionality was preserved, the HTML was simplified and modified for appropriate consumption and interaction on the PDA form factor, as can be seen in Figure 12.

Although SALT technology can be demonstrated using a Pocket PC PDA, the speech processing is not performed locally on the mobile device but redirected to a server machine on the LAN. While this is not practical for deployment, it enabled us to experiment with the approach. It is anticipated that speech processing on the PDA will become available in the near future.

Figure 10. While taking measurements from the patient, the nurse can speak the clinical measurements via the Bluetooth enabled headset directly into the HTML form

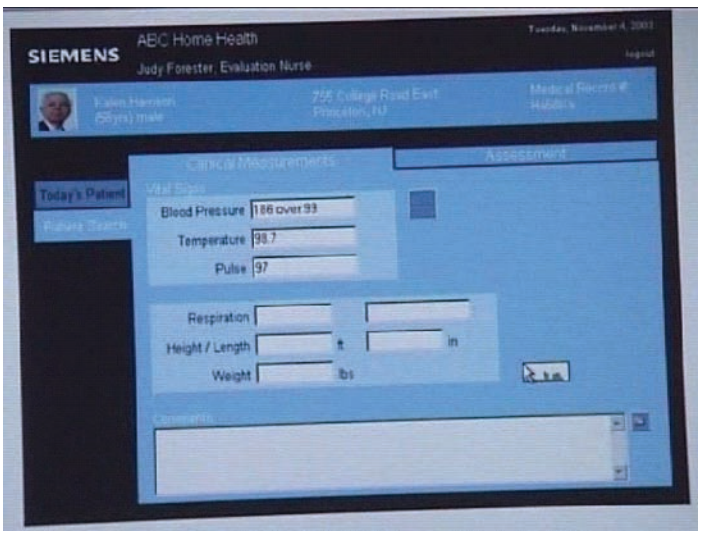


Figure 11. While interviewing the patient, the nurse can speak the diagnosis and functional assessment data via the Bluetooth enabled headset directly into the HTML form

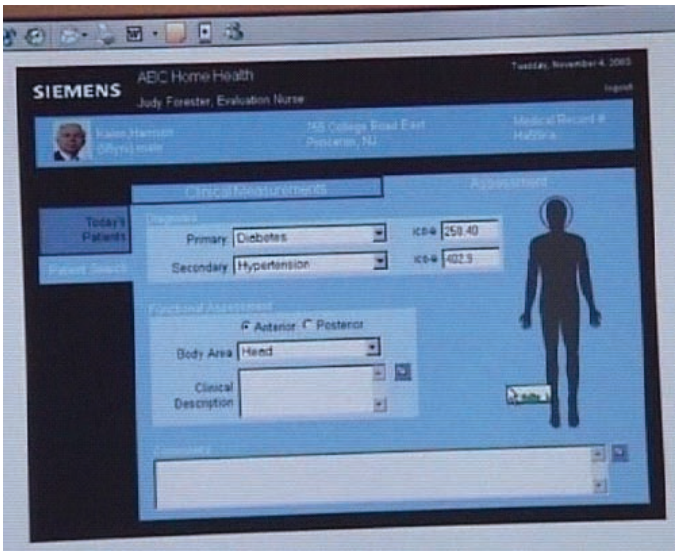
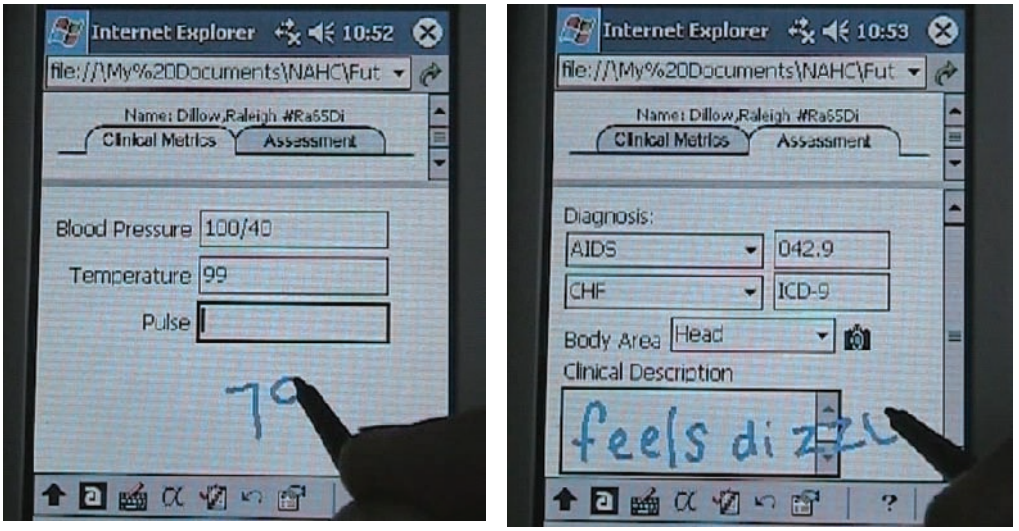


Figure 12. Leveraging handwriting recognition to capture vital signs into an HTML form

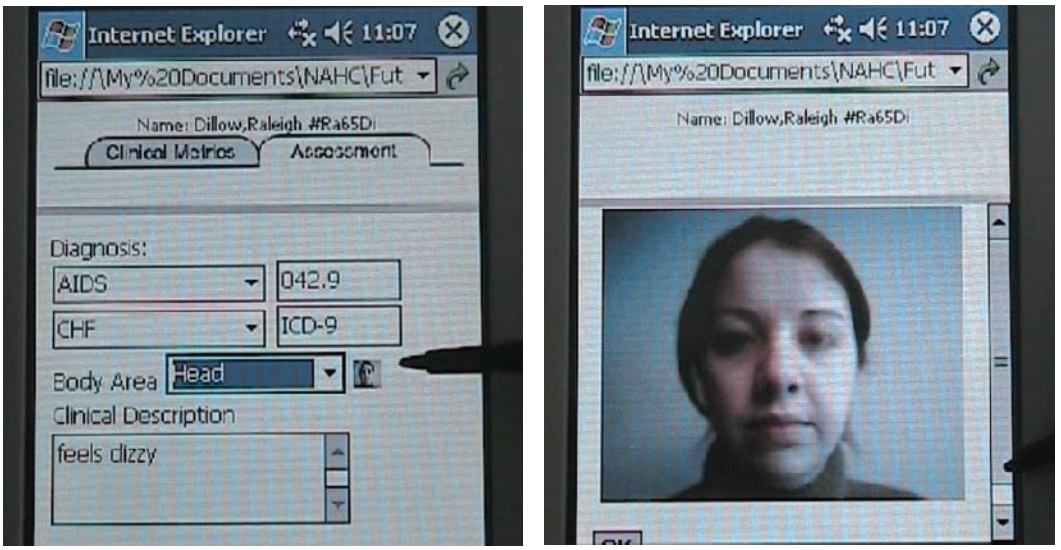


Speech is not the most appropriate input mechanism for every occasion, but one tool in the palette of a multimodal interface designer. As many of the nurses that we studied rely on paper and pencil to record patient notes, we sought to exploit handwriting recognition technology as a means to harness this activity and increase nurses' productivity by capturing this text and entering in directly into HTML form fields. This approach can be seen in Figure 12.

In our interviews we found that the old adage of a picture capturing a thousand words is not lost on nurses. Hence, we sought to offer in our prototype the seamless support for capturing

Figure 13. Convenient image capture is seamlessly integrated into the patient record

Figure 14. A thumbnail icon of the captured image can simply be clicked-on for closer inspection



images and integrating them directly into the patient information. A small and inexpensive digital camera peripheral connected via the SDIO port can be used for this purpose. The process of capture, abstraction through a thumbnail representation, and the viewing of the image can be seen in Figures 13 and 14.

In addition to capturing an image, it is often desirable to be able to augment an image with a synchronized audio commentary and associated pen markings to describe specific aspects of the patient's condition. Hence, a multimedia annotation capability is offered that can support this requirement and be stored along with the patient record.

Current and future work includes conducting formal user studies with nursing practitioners from two home healthcare agencies to evaluate the usability and efficacy of the devices and approaches described above to inform the subsequent iteration of the prototype design.

CONCLUSIONS

The manner in which a networked enterprise can facilitate a plethora of ways for healthcare personnel to communicate and collaborate was explored within this chapter. The next generation of communication technologies augurs well for converged voice and data solutions on a single network. We anticipate a closer union between healthcare IT systems and web-based communications. Technology innovation has spawned a proliferation of communication and data devices, such as GPRS cellular phones and PDAs, and with it a significant opportunity for accessing clinical information anywhere, anytime, allowing healthcare practitioners to collaborate upon clinical information and reach conclusions more effectively.

Throughout the chapter a selection of technologies has been presented that can enable healthcare personnel to interact in a variety of styles. These approaches demonstrated how multimedia technologies can be harnessed to capture information and enable users to collaborate in range of modes in an effective manner. It was also described how the knowledge captured during collaboration interactions can be saved and archived for future search, reference and reusability. In addition, technology support for home healthcare nurses is presented that shows potential for streamlining the capture and entry of information into the patient record.

ACKNOWLEDGMENTS

The authors would like to thank our team members at Siemens Corporate Research for their contributions to the technologies described.

REFERENCES

Bluetooth (n.d.). Online http://www.bluetooth.com

Bluetooth headset from Siemens. Online http://www.siemens-mobile.com/cds/frontdoor/0,2241,hq_en_0_49735_rArNrNrNrN,00.html

Giles, T. (1996). The cost-effective way forward for the management of the patient with heart failure. *Cardiology, 87*(1), 33-39.

Hibbert, D., Mair, F.S., May, C.R., Boland, A., O'Connor, J., Capewell, S., & Angus, R.M. (2004). Health professionals responses to the introduction of a home telehealth service. *Journal of Telemedicine and Telecare 10*(4), 226-230.

HIPPA (n.d.). Online http://www.hhs.gov/ocr/hipaa/

Multimodal Interaction Working Group (n.d.). Online http://www.w3.org/2002/mmi/

SALT Forum (n.d.). Online http://www.saltforum.org

Sastry, C., Lewis, D. & Pizano, A. (1999). Webtour: A system to record and playback dynamic multimedia annotations on Web document content. *Proceedings of the ACM International Conference on Multimedia*, Orlando, October (pp. 175-178).

Siemens Biometric Mouse (n.d.). Online http://www.siemensidmouse.com/

Siemens IMS (n.d.). Online http://www.siemens-mobile.com/cds/frontdoor/0%2C2241%2Chq_en_0_860_rArNrNrNrN%2C00.html

Siemens OpenScape (n.d.). Online http://www.siemensenterprise.com/prod_sol_serv/products/openscape/

Siemens Soarian (n.d.). Online http://www.medical.siemens.com/webapp/wcs/stores/servlet/CategoryDisplay?storeId=10001&langId=-1&catalogId=-1&catTree=100001,19051,19027&categoryId=19027

Siemens Information and Communications Mobile (n.d.). Online http://www.siemens-mobile.com

Siemens Information and Communications Networks (n.d.). Online http://www.icn.siemens.com

Siemens Medical (n.d.). Online http://www.medical.siemens.com

This work was previously published in Clinical Knowledge Management: Opportunities and Challenges, edited by B. Bali, pp. 139-158, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.22
Multimedia Systems and Content-Based Image Retrieval

Sagarmay Deb
University of Southern Queensland, Australia

ABSTRACT

In this chapter, we present a basic introduction of the two very important areas of research in the domain of information technology, namely, multimedia systems and content-based image retrieval. The latter is still a widely unresolved problem. We discuss some of the works done so far in content-based image retrieval in the context of multimedia systems.

INTRODUCTION

Research on multimedia systems and content-based image retrieval has gained tremendous momentum during the last decade. This is due to the fact that multimedia databases cover the text, audio, video and image data which help us to receive enormous amounts of information and which has brought fundamental changes in our life style. Content-based image retrieval (CBIR) is a bottleneck of the access of multimedia databases simply because there remains vast differences in the perception capacity between a human being and a computer. We give an introduction of these two areas in this chapter along with an idea of the research being conducted in CBIR over the last few years in the context of multimedia systems. The section on multimedia databases describes general background of multimedia databases. The section on Content-based image retrieval talks about general background, problems and some of the works done so far in content-based image retrieval. A summary is presented at the end.

MULTIMEDIA DATABASES

General Background

Multimedia environment, much talked-about topic in computers, is adding new dimensions in the area of information technology and all relevant fields. Multimedia and imaging applications are enriching existing applications by integrating images, voices, text and video. The ultimate goal is to have all information types digitized and computerized. As we proceed to convert enormous amounts of non-digitized information into multimedia

objects, the outcomes are extremely significant with its effects in various activities of life.

The available literature talks about certain advanced technologies which are:

- Computer-aided design and manufacturing (CAD/CAM) using 2-D and 3-D graphics, animation and visualization
- Network standard, such as asynchronous transfer mode (ATM) and fibre distributed data interface (FDDI) which allow multimedia information to flow through wide area networks (WANS) as well as local area networks (LANS)
- Image processing systems which include more sophisticated techniques to segment and process images
- Character and object recognition systems using pattern recognition and neural networks
- Software systems such as object-oriented languages, databases and operating systems
- Hardware components such as video cameras, video boards, sound boards and high-resolution monitors
- Storage devices such as optical disks and magneto-optic technologies

All of the above developments in information technology are lending support to multimedia information technology.

Multimedia environments can be available in PC-based systems, then in more advanced Power PCs with built-in audiovisual capabilities and to more advanced and sophisticated UNIX workstations.

There are some similarities between object-oriented databases and multimedia databases. Just as an object-oriented database tries to depict the real world as much as possible so are multimedia databases because of their ability to capture and play back the sounds and images of the real world as realistically as possible.

Other features of multimedia databases are:

1. Automatic feature extraction and indexing where image processing and pattern recognition systems for images often extract the various features and content of multimedia objects. When large amounts of non-digitized information are converted into digital multimedia objects, quite a few automatic indexing features are utilized. This is particularly true where paper documents are converted into scanned digitized images that are eventually recognized by optical character recognition (OCR) products.
2. Interactive querying, relevance feedback and refining where user queries involve a graphical user interface where the construction of the query can be performed interactively. In multimedia databases, it is common to have domains or lists of various existing multimedia objects so that the user can construct a query interactively. The result of the query can be shown with more relevant selections first followed by less relevant selections and the user can select relevant objects, refine the query and resubmit, etc.
3. Content-based indexing where a video could contain the frame number of the start of each clip or scene for each video. For images containing objects arranged spatially, iconic indexing through 2-D strings contains the location and relative position of each element or object in the whole image.
4. Single and multikey (dimensional) indexing where spatial and multidimensional indexes are used. A geometric region has X and Y coordinates. Spatial indexes, such as R trees or multidimensional structures such as grid files or k-d trees will produce better access times for queries covering all the dimensions of the multidimensional attribute being retrieved.
5. Storage organization for Binary Large Objects (BLOB) where multimedia databases

allow the users to create and store BLOBs. Because the BLOB interfaces typically allow the user to access and update byte or bit streams, the multimedia storage manager makes use of the positional indexes to store the BLOBs. This permits very fast access to continuous streams of bytes or bits starting at a particular position. Positional indexes also accelerate the insertions and appends of byte streams in BLOBs.

6. Spatial data types and queries where the spatial relationships between various elements of the image are studied.
7. Query optimization where a multimedia database management system that integrates hierarchical storage optimization, information/content-retrieval modules and complex object managers must decide on the best way to perform execution.
8. Complex object clustering where to provide efficient access to a multimedia complex object, the storage management layer has to incorporate complex object clustering storage techniques. The goal would be to minimize I/O access time for the complex object and its sub-objects and also the processing time needed to reconstruct the complex object.

Multimedia Database Management Systems

Multimedia Database Management Systems (MDBMS) deal with audio, video, text and image data. We can create, update, delete and enquire these data in the multimedia databases. Ordinarily, data are BLOBS. Multimedia objects can take lot of storage space and that requires very efficient storage management systems for fast and efficient access of the data.

There are three layers in a multimedia database. The interface layer deals with browsing and query. Then, the object composition layer handles managing objects and the storage layer deals with clustering and indexing.

Research Issues

Some of the important research problems are: (1) watermarking technology, which is very useful in protecting digital data such as audio, video, image, formatted documents and three dimensional objects; (2) synchronization and timeliness, which are required to synchronize multiple resources like audio and video data; (3) quality of service (QoS), which is relevant to high-speed networks to achieve low end-to-end delay, loss rate and optimum data transmission with the available resources; and (4) reusability, where browsing of objects give the users the facility to reuse multimedia resources. (5) Since the multimedia data is voluminous and occupy a lot of storage space, developing efficient storage techniques is essential for fast accessing and retrieval of information. (6) For video content-based retrieval, efficient shot boundary detection, key frames selection, feature extraction and retrieval are important research issues at the moment.

CONTENT-BASED IMAGE RETRIEVAL

General Background

Images are being generated at an ever-increasing rate by various sources. They include military purposes, aviation satellites, biomedical purposes, scientific experiments and home entertainment.

Application areas in which CBIR is a principal activity are numerous and diverse. They are:

- Art galleries and museum management
- Architectural and engineering design
- Interior design
- Remote sensing and management of earth resources
- Geographic information systems
- Scientific database management
- Weather forecasting

- Retailing
- Fabric and fashion design
- Trademark and copyright database management
- Law enforcement and criminal investigation
- Picture archiving and communication systems (Gudivada & Raghavan, 1995)

Previously there were two approaches to CBIR.

The first one is the attribute-based representation advocated by database researchers where image contents are modeled as a set of attributes extracted manually and managed within the framework of conventional database management systems. Queries are specified using these attributes. This entails a high-level of image abstraction.

The second approach propagated by image interpretation researchers depends on an integrated feature-extraction/object-recognition subsystem to overcome the limitations of attribute-based retrieval. This subsystem automates the feature-extraction and object-recognition task that occurs when the image is inserted into the database. This automated approach to object recognition is computationally expensive, difficult and tends to be domain specific.

Recent CBIR research tries to combine both of the above mentioned approach and has given rise to efficient image representations and data models, query-processing algorithms, intelligent query interfaces and domain-independent system architecture.

There are two major categories of features. One is primitive, which is concerned with extracting boundaries of the image and the other one is logical, which defines the image at various levels of detail.

Regardless of which approach is used, the retrieval in CBIR is done by color, texture, sketch, shape, volume, spatial constraints, browsing, objective attributes, subjective attributes, motion, text and domain concepts (Gudivada & Raghavan, 1995).

Our study of the existing literature suggests, in recent times, there have been very many attempts to perform CBIR on an efficient basis based on feature, color, texture and spatial relations. Quite a few models have been developed which address the problem of image retrieval from various angles. Some of the models are QBIC (Query by Image Content), Virage, Pichunter, VisualSEEK, Chabot, Excalibur, Photobook, Jacob, and Digital Library Project (Seaborn, 1997).

One of the most well known packages is QBIC which was developed by IBM. It uses color, texture, shape, example images and sketches to retrieve images used in large image and video databases. Based on the same kind of approach developed by Virage which made use of color, composition (color layout), texture and structure (object boundary information) and is being applied in face recognition and in retrieval of ophthalmologic images. NEC Research Institute has developed Pichunter, which utilizes image properties like ratio of image dimensions, color percentages and global statistical and frequency properties. It is more applicable in database retrieval instead of feature detection and is also being applied in relevance feedback using Bayesian probability theory. VisualSEEK, developed at the Columbia University Centre for Telecommunication Research, uses color percentage method in content-based retrieval. Using regional colors and their relative locations, the image is segmented and this is quite a bit similar to the way we perceive an image. Chabot mainly uses texts to retrieve images. It uses to some extent color percentages to retrieve images automatically otherwise all features are input manually. Excalibur is of the same type as QBIC and Virage and uses standard metrics, color, shape and texture and like Pichunter uses image ratio. In addition, it extracts features like structure of brightness and color. It gives an option to the users to indicate which features are dominant. MIT's Vision and

Modeling Group developed Photobook, which uses color percentages, textures and statistical analysis. The images are segmented here and then texts are made out of those segmentations using predefined templates and techniques. The images are retrieved based on these texts. Jacob uses a combination of color, texture and motion as features to retrieve video clips as this package is developed for video databases. The Digital Library Project of Berkeley University applies color percentage method and feature of dots. The user can define the quantity of various colors in the image and also can define colors and sizes of dots to be there in the image.

These models have brought this area of CBIR from its infancy to a matured stage. They study the various features of the images, make statistical analysis of color distribution as well as shape and texture and retrieve images from the contents of the image.

The developments in this field have been defined in three levels (Eakins & Graham, 1999).

Level one is the primitive level where low-level features like color, texture, shape and spatial locations are used to segment images in image databases and then find symmetry based on these segmentations with the input image. Plenty of research projects were being done during the past decade. As we mentioned earlier, many software packages have been developed for efficient image retrieval. Most of them have used a combination of text-based and content-based retrieval. Images are segmented manually, a priory and text are generated based on these manual segmentations and retrievals are conducted accordingly. But since the volume of images generated could be enormous in fields like satellite picturing, this method of manual part processing is time-consuming and expensive. Few automatic retrievals without human intervention have been developed like QBIC, Virage, and Excalibur which are now commercially being used in addition to packages developed which are not yet commercial. But they have limited applications in areas like trademark registration, identification of drawings in a design archive or color matching of fashion accessories based on input image. No universally accepted retrieval technique has yet been developed. Also, the retrieval techniques developed without human intervention are far from perfect. Segmentation has been done in some packages based on color where the segmented parts taken individually do not contribute to any meaningful identification (Ma et al., n.d.). They generate a vague symmetry between input objects and objects in image database. This level still needs to be developed further to have universal applications.

Level two deals with bringing out semantic meanings of an image of the database. One of the best known works in this field is of Forsyth et al. (1996) successfully identifying human beings within images. This technique had been applied for other objects like horses and trees.

Also, for example, a beach can be identified if a search is based on color and texture matching and the color selected is wide blue with yellow texture below.

Attrasoft Image Finder has developed an image retrieval technique where input images would be stored in various files. Also, images would be kept in directory files. There is an interface screen where users can provide the file name containing the input image and also can put various parameters like focus, background, symmetry, rotation type, reduction type and so on. The images from the directory would then be retrieved based on these inputs. The images in the directory are defined containing the sample segments or translated segments, rotated segments, scaled segments, rotated and scaled segments, and brighter or darker segments. This method goes to some extent in bringing out semantic meanings in an image in the sense that the user can specify an input image semantically, then corresponding input image is retrieved and based on that input image, image database is searched to find symmetry (Attrasoft, 2001). There are few other similar automatic image retrieval models available including Computer Vision Online Demos.

But this level also needs a lot more developments to achieve a universally accepted technique to bring out semantic meanings out of the image.

Level three attempts retrievals with abstract attributes. This level of retrieval can be divided into two groups. One is a particular event such as 'Find pictures of Australia playing cricket against another particular country.' A second one could be 'Find a picture which is a residential area.'

To interpret an image after segmentations and analyzing it efficiently requires very complex reasoning. This also requires a retrieval technique of level two to get semantic meanings of various objects. It is obvious this retrieval technique is far from being developed with modern technology available in the field.

Works Done So Far

In this section, we present some of the works done in the field of content-based image retrieval. Although plenty of research works have been done so far in the field, no universally accepted model has yet been developed. The research concentrated on image segmentation based on low-level features like color, shape, texture and spatial relations. But to find the semantic meanings or high-level meanings of an image like whether it is the image of human beings or a bus or a train and so on is still a problem. Attempts are being made to link low-level and high-level features. But it is proving difficult for the very simple reason there remains a vast gap between human perception and computer perception. We present a few references of the works done.

One approach is where document images are accessed directly, using image and object attributes and the relative positions of objects within images, as well as indirectly, through associated document components. This is based on retrieving multimedia documents by pictorial content. Queries may address, directly or indirectly, one or more components. Indirect addressing involves references from associated components, e.g., an image caption is a text component referring to an image component and so is an in-text reference to an image. A symbolic image consists, in general, of objects, relations among objects and descriptions of object and image properties. The properties of objects and whole images are described by object and image attributes, respectively (Constantopoulos et al., 1991).

Then there is another method of querying and content-based retrieval that considers audio or visual properties of multimedia data through the use of MORE. In MORE, every entity in an application domain is represented as an object. An object's behavior is presented with a method, which activates the object by receiving a certain message. Objects bearing the same characteristics are managed as a single class. Generally, a class contains structural definitions, methods, or values that the objects in the class commonly posses (Yoshitaka et al., 1994).

Range searches in multi-dimensional space have been studied extensively and several excellent search structures have been devised. However, all of these require that the ranges in the different dimensions be specified independently. In other words, only rectangular regions can be specified and searched so far. Similarly, non-point objects can be indexed only in terms of their bounding rectangles. A polyhedral search of regions and polyhedral bounding rectangles can often provide a much greater selectivity in the search. How to use multi-attribute search structures for polyhedral regions, by mapping polyhedral regions into rectangular regions of a higher dimensions has been shown (Jagadish, 1990).

In human design processes, many drawings of shapes remain incomplete or are executed inaccurately. Cognitively a designer is able to discern these anomalous shapes, whereas current CAAD systems fail to recognize them properly so that CAAD systems are unable to match left-hand-side conditions of shape rules. As a result, current CAAD systems fail to retrieve

right-hand-side actions. Multi-layered neural networks are constructed to solve the recognition and transformation of ill-processed shapes in light of recent advances of connectionism in cognitive psychology and artificial intelligence (Liu, 1993).

Attempts have been made to retrieve a similar shape when shapes are measured by coordinate systems (Mehrotra & Gary, 1995).

Then, MIRO (Multimedia Information Retrieval) is researching new methods and techniques for information storage and retrieval so that all types of media can be handled in an integrated manner through adaptive interaction with a user. Topics include human-computer interaction, logical and probabilistic models of information retrieval, run-time support for multimedia information retrieval and evaluation of retrieval effectiveness (Thanos, 1995).

A parallel computing approach to creating engineering concept spaces for semantic retrieval has been developed through the Illinois Digital Library Initiative Project. This research presents preliminary results generated from the semantic retrieval research component of the Illinois Digital Library Initiative (DLI) project. Using a variation of the automatic thesaurus generation techniques, to which is referred to as the concept space approach, it is aimed to create graphs of domain-specific concepts (terms) and their weighted co-occurrence relationships for all major engineering domains. Merging these concept spaces and providing traversal paths across different concept spaces could potentially help alleviate the vocabulary (difference) problem evident in large-scale information retrieval. Experiments had been done previously with such a technique for a smaller molecular biology domain (Worm Community System, with 10+ MBs of document collection) with encouraging results (Chen et al., 1996).

A system named MARCO (denoting MAp Retrieval by COntent) that is used for the acquisition, storage, indexing and retrieval of map images is presented. The input to MARCO is raster images of separate map layers and raster images of map composites. A legend-driven map interpretation system converts map layer images from their physical representation to their logical representation. This logical representation is then used to automatically index both the composite and the layer images. Methods for incorporating logical and physical layer images as well as composite images into the framework of a relational database management system are described. Indices are constructed on both the contextual and the spatial data thereby enabling efficient retrieval of layer and composite images based on contextual as well as spatial specifications. Example queries and query processing strategies using these indices are described. The user interface is demonstrated via the execution of an example query. Results of an experimental study on a large amount of data are preserved. The system is evaluated in terms of accuracy and in terms of query execution time (Samet, 1996).

Fingerprint databases are characterized by their large size as well as noisy and distorted query images. Distortions are very common in fingerprint images due to elasticity of the skin. In this article, a method of indexing large fingerprint image databases is presented. The approach integrates a number of domain-specific high-level features such as pattern class and ridge-density at higher levels of the search. At the lowest level, it incorporates elastic structural feature-based matching for indexing the database. With a multilevel indexing approach, the search space is reduced. The search engine has also been implemented on Splash 2 — a field programmable gate array (FPGA)-based array processes to obtain near ASIC level speed of matching. This approach has been tested on a locally collected test data and on NIST-S, a large fingerprint database available in the public domain (Ratha et al., 1996).

Focussing has been done on the use of motion analysis to create visual representations of videos that may be useful for efficient browsing and in-

dexing in contrast with traditional frame-oriented representations. Two major approaches for motion based representations have been presented. The first approach demonstrated that dominant 2-D and 3-D motion techniques are useful in their own right for computing video mosaics through the computation of dominant scene motion and/or structure. However, this may not be adequate if object level indexing and manipulation is to be accomplished efficiently. The second approach presented addresses this issue through simultaneous estimation of an adequate number of simple 2-D motion models. A unified view of the two approaches naturally follows from the multiple model approach: the dominant motion method becomes a particular case of the multiple motion method if the number of models is fixed to be one and only the robust EM algorithm without the MDL stage employed (Sawhney, 1996).

Image content-based retrieval is emerging as an important research area with application to digital libraries and multimedia databases. The focus is being put on the image processing aspects and, in particular, using texture information for browsing and retrieval of large image data. It proposed use of Gabber wavelet features for texture analysis and provided a comprehensive experimental evaluation. Comparisons with other multi-resolution texture features using the Brodatz texture database indicate that the Gabor features provide the best pattern retrieval accuracy. An application to browsing large air photos is illustrated (Manjunath & Ma, 1996).

A method is developed for the content-based retrieval of multi-spectral satellite images using invariant representations. Since these images contain a wide variety of structures with different physical characteristics it is useful to exploit several classes of representations and algorithms. Working from a physical model for multi-band satellite image formation, existing algorithms for this application have been modified and integrated. The performance of the strategy has been demonstrated for image retrieval invariant to atmospheric and illumination auditions from a database of 166 multi-band images acquired at different times over areas of the U.S. (Healey & Jain, 1996).

The problem of retrieving images from a large database is addressed using an image as a query. The method is specifically aimed at databases that store images in JPEG format and works in the compressed domain to create index keys. A key is generated for each image in the database and is matched with the key generated for the query image. The keys are independent of the size of the image. Images that have similar keys are assumed to be similar, but there is no semantic meaning to the similarity (Shneier & Abdel-Mottaleb, 1996).

The Multi-mission VICAR Planner (MVP) is described in an article. It is an AI planning system which uses knowledge about image processing steps and their requirements to construct executable image processing scripts to support high-level science requests made to the Jet Propulsion Laboratory (JPL) Multi-mission Image Processing Subsystem (MIPS). This article describes a general AI planning approach to automation and application of the approach to a specific area of image processing for planetary science applications involving radiometric correction, color triplet reconstruction and mosaicing in which the MVP system significantly reduces the amount of effort required by image processing experts to fill a typical request (Chien & Mortensen, 1996).

Another paper presents how morphological transformations can be related to representations of a set on different lattices. (A hierarchical definition of structuring element conveys to a class of multi-grid transformations Pk that handle changes on discrete representations of regions.) The transformations correspond to upward and downward processes in a hierarchical structure, based on multi-grid transformations, a method to delineate non-perfectly-isolated objects in an nxn image requiring O(log n)time is presented. The approach considers grey level regions as sets

and processes through a pyramid to carry out geometric manipulations. Extending the concept of boundary to cope with hierarchical representations of a set, a second method, which identifies the boundaries in an image is discussed (Montiel et al., 1996).

IMEDIA project, which is related to image analysis, is the bottleneck of multimedia indexing concerns about image analysis for feature space and probabilistic modelisation, statistics and information theory for interactive browsing, similarity measure and matching. To achieve these goals, research involves the following topics: image indexing, partial queries, interactive search, multimedia indexing (Boujemma et al., 2000).

In a project named Efficient Content-Based Image Retrieval, the focus is the development of a general, scalable architecture to support fast querying of very large image databases with user-specified distance measures. They have developed algorithms and data structures for efficient image retrieval from large databases with multiple distance measures. They are investigating methods for merging their general distance-measure independent method with other useful techniques that may be distance measure specific, such as keyword retrieval and relational indexing. They are developing both new methods for combining distance measures and a framework in which users can specify their queries without detailed knowledge of the underlying metrics. They have built a prototype system to test their methods and evaluated it on both a large general image database and a smaller controlled database (Shapiro et al., 2000).

There is another work that addresses the issue of a gap existing between low-level visual features addressing the more detailed perceptual aspects and high-level semantic features underlying the more general aspects of visual data. Although plenty of research works have been devoted to this problem, so far, the gap still remains (Zhao et al., 2002).

Another chapter provides a state-of-the-art account of Visual Information Retrieval (VIR) systems and Content-Based Visual Information Retrieval (CBVIR) systems. It provides directions for future research by discussing major concepts, system design issues, research prototypes and currently available commercial solutions (Marques et al., 2002).

SUMMARY

A general introduction of the subject-area of the book has been given in this chapter. Both multimedia systems and content-based image retrieval have been discussed from the introductory level. Also, focus has been made to specify current problems in both of these fields and research efforts being made to solve them. Some of the research works done in the field of content-based image retrieval have been presented to give an idea of the research being conducted. All these should give a broad picture of the contents of issues, covered in these areas.

REFERENCES

Attrasoft (2001, May). *Attrasoft*, P.O. Box 13051, Savannah, GA. 31406, USA.

Boujemma, N. et al. (2000, February). *IMEDA Project*, INRIA.

Chen, H., Scatz, B., Ng, T., Kirchhoff, A., & Lin, C. (1996). A parallel computing approach to creating engineering concept spaces for semantic retrieval: The Illinois digital library initiative project. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (18).

Chien, S.A. & Mortensen, H.B. (1996). Automating image processing for scientific data analysis of a large image database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8).

Constantopoulos, P., Drakopoulos, J., & Yeorgaroudakis, Y. (1991). Retrieval of multimedia documents by pictorial content: A prototype system. *International Conference on Multimedia Information Systems '91.* The Institute of Systems Science, National University of Singapore, 35-48.

Eakins, J.P. & Graham, M.E. (1999). Content-based image retrieval: A report to the JISC technology application programme. Institute for Image Data Research, University of Northumbria at Newcastle, UK.

Forsyth, D.A et al. (1996). Finding pictures of objects in large collections of images. In Heidon, P.B & Sandore, B. (Eds.), *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing*, (pp. 118-139). Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Gudivada, V.N. & Raghavan, V.V. (1995). Content-based image retrieval systems. *IEEE,* September 1995.

Healey, G. & Jain, A. (1996). Retrieving multispectral satellite images using physics-based invariant representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8).

Jagadish, H. V. (1990). Spatial search with polyhedra. *IEEE,* 311-319.

Liu, Y. (1993). A connectionist approach to shape recognition and transformation. *CAAD Futures '93,* 19-36.

Ma, W.Y., Manjunath, B.S., Luo, Y., Deng, Y., & Sun, X. (n.d.). *NETRA: A Content-Based Image Retrieval System.* Dept. of Electrical and Computer Engineering, University of California, Santa Barbara, California, USA.

Manjunath, B.S. & Ma, W.Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(8), August 1996.

Marques, O. & Furht, B. (2001). Content-based visual information retrieval. *Distributed Multimedia Databases: Techniques and Applications,* 37-57.

Mehrotra & Gary. (1995). Similar-shape retrieval in shape data management. *IEEE,* September 1995.

Montiel, M.E., Aguado, A.S., Garza-Jinich, M.A., & Alarcdn, J. (1996). Image manipulation using m-filters in a pyramidal computer model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8).*

Ratha, N.K., Karu, K., Chen, S., & Jain, A.K. (1996). A real-time matching system for large fingerprint databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8).*

Samet, H. (1996). MARCO: Map retrieval by content. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8).*

Sawhney, H. & Ayer, S. (1996). Compact representations of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18(8).

Seaborn, M.A. (1997). Image Retrieval Systems.

Shapiro, L.G. et al. (2000). *Efficient Content-Based Image Retrieval.* Department of Computer Science and Engineering, University of Washington.

Shneier, M. & Abde-Mottaleb, M. (1996). Exploiting the jpeg compression scheme for image retrieval. *IEEE Transactions on pattern Analysis and machine Intelligence,* 18(8).

Thanos, C. (1995). Multimedia information retrieval. *Sven MiiBig.*

Yoshitaka, A., Kishida, S., & Hirakawa, M. (1994). Knowledge-assisted content-based retrieval for multimedia databases. *IEEE Multimedia,* (Winter), 12-21.

Zhao, R. & Grosky, W.I. (2001). Bridging the semantic gap in image retrieval. *Distributed Multimedia Databases: Techniques and Applications*, 14-36.

This work was previously published in Multimedia Systems and Content-Based Image Retrieval, edited by JS. Deb, pp. 1-13, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.23
Multimedia Digital Library as Intellectual Property

Hideyasu Sasaki
Keio University, Japan

Yasushi Kiyoki
Keio University, Japan

ABSTRACT

The principal concern of this chapter is to provide those in the digital library community with the fundamental knowledge on the intellectual property rights and copyrights regarding multimedia digital libraries. The main objects of our discussion are the multimedia digital libraries with content-based retrieval mechanisms. Intellectual property rights are the only means for database designers to acquire their incentive of content collection and system implementation in database assembling. We outline the legal issues on multimedia digital libraries and retrieval mechanisms. As the protection of intellectual property rights is a critical issue in the digital library community, the authors present legal schemes for protecting multimedia digital libraries and retrieval mechanisms in a systematic, engineering manner.

INTRODUCTION

Digital library is the global information infrastructure in the networked society (Borgman, 2000). The rights protection of multimedia digital libraries and retrieval mechanisms is a critical issue in the digital library community that demands intellectual property schemes for recouping their investment in database design and system implementation. In this chapter, we describe the technical and legal issues on multimedia digital libraries and retrieval mechanisms as intellectual properties.

The purpose of this chapter is to discuss copyright and intellectual property rights on digital libraries from the designer/architecture perspective, which has not been discussed with sufficient attention at the present. The end-user perspective has been discussed as an important

element for users of information services, including librarians, in the context of copyright law on multimedia digital libraries. Its typical case is the public use of copyright for educational or academic service. Content creators of digital libraries have definitely enjoyed copyright enforcement over their works under that current legal scheme. However, the designers or architectures of multimedia digital libraries do not have proper foundations for their rights protection that is to be equivalent to the copyright protection. Under this designer/architecture perspective, we especially focus on content-based retrieval and its application to multimedia digital libraries. Content-based retrieval is a promising technique for networked multimedia digital libraries whose tremendous volume demands automatic indexing rather than manual indexing for retrieval operations.

The scope of this chapter is also restricted within the current standard of laws and cases for transnational transaction and licensing of digital copyright and intellectual property rights regarding multimedia digital libraries. Cultural diversity in the Asia-Pacific region allows a number of legislative differences in copyright and intellectual property laws. Meanwhile, digital content is the object of its worldwide transaction. The harmonization of its related rights is inevitable because a number of countries have joined international trade agreements on intellectual property rights. We need a clear and uniform standard with which the Asian-Pacific countries are able to keep up with the foregoing countries.

In this chapter, we discuss three current issues on multimedia digital libraries and intellectual property laws, and then present three types of intellectual property schemes, respectively. The first issue is copyright protection of indexed digital contents that are stored in digital libraries. Its corresponding scheme is for copyrighting multimedia digital libraries that are associated with keyword-based retrieval operations. The second issue is patentability of retrieval mechanisms. Its corresponding scheme is for patenting content-based retrieval processes in multimedia digital libraries. Finally, the last issue is the limitations of copyright in the advent of content-based retrieval. Its corresponding scheme is a promising direction, which leverages the *de facto* protection of multimedia digital libraries that are associated with content-based retrieval operations.

BACKGROUND

Intellectual property law gives incentive to advance appropriate investment in database design and implementation (Jakes & Yoches, 1989). However, present legal studies are not satisfactory as the source of technical interpretation of the intellectual properties regarding multimedia digital libraries and retrieval mechanisms. With the advent of content-based image retrieval (CBIR), we face novel issues on the right protection of multimedia digital libraries and retrieval mechanisms.

Multimedia digital libraries consist of digital contents and retrieval systems. The digital contents are copyrightable materials whose stakeholders are content creators. The retrieval systems are patentable processes of database designers.

The principal problem discussed in this chapter is restricted into the conflicting interests between content creators and database designers. Intellectual property lawyers in the area of digital libraries have somehow neglected this designer/architecture perspective. The issues related to the problem are found in three stages. First, copyright does not always solve conflicting interests between content creators and database designers. Content creators create copyrightable individual contents, for example, pictures and images. Database designers integrate the entire content of each multimedia digital library with indexes or metadata. The problem is deciding which component differentiates the entire content of each digital library as an independent object of right protection from its copyrightable individual

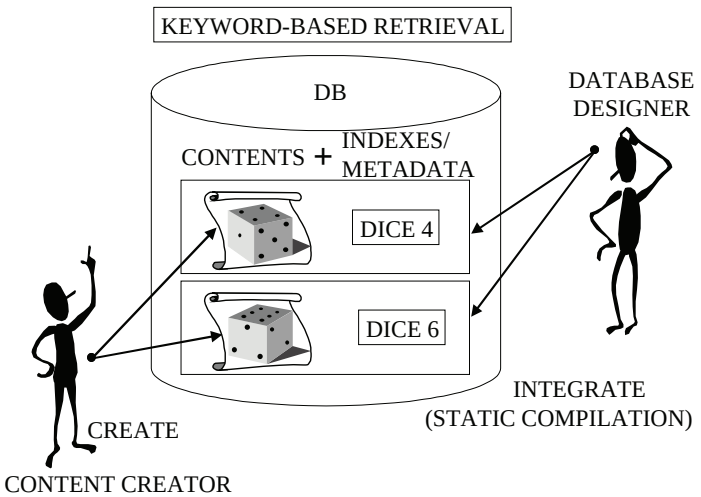
contents. Second, CBIR mechanisms need the specific examination of patentability. The computer-implemented processes in CBIR consist of the combinations of means, some of which are prior disclosed inventions. Those processes consist of the means for parameter setting that is adjusted to retrieve specific kinds of images in certain narrow domains. The problem is determining which process is of technical advancement (non-obviousness) based on its combinations of the prior arts and parameter setting. Finally, the dynamically indexed content of multimedia digital libraries goes beyond the scope of conventional copyright protection. CBIR mechanisms generate indexes to individual contents every time queries are requested for retrieval. The problem is selecting which component leverages the protection of each multimedia digital library, as the entire content is independent of its individual contents.

INDEXES AND COPYRIGHT PROTECTION

U.S. Copyright Act (1976) defines that a compilation or assembling of individual contents, that is, preexisting materials or data, is a copyrightable entity as an original work of authorship. Typical examples of copyrightable multimedia materials stored in digital libraries take a variety of forms like text or images, photos or video streams. Figure 1 outlines that the collection of static indexes and individual contents forms a component of contents-plus-indexes. That component identifies the entire content of each database, as is a static and copyrightable compilation. Indexes including metadata must be statically assigned to individual contents that are restored in databases for keyword-based retrieval operations.

Gorman & Ginsburg (1993) and Nimmer et al. (1991) state that a compilation is copyrightable as far as it is an "original work of authorship that is fixed in tangible form." A database is copyrightable in the form of a component of contents-plus-indexes while static indexes or metadata are fixed in a tangible medium of repository, database. In keyword-based retrieval, static indexes or metadata represent a certain kind of categorization of the entire content of each database. The originality on the categorization makes each database copyrightable as is different from the mere collection of its individual contents. What kind of categorization should be original to constitute a copyrightable compilation? The case

Figure 1. Digital library of visual content associated with keyword-based retrieval



of American Dental Association v. Delta Dental Plan Association (1997) determined that minimal creativity in compilation sufficed this requirement of originality on databases. A uniform scheme on the categorization regarding indexes or metadata must be formulated in an engineering manner.

COPYRIGHTING MULTIMEDIA DIGITAL LIBRARIES WITH KEYWORD-BASED RETRIEVAL

We technically interpret a scheme for copyrighting the multimedia digital libraries with keyword-based retrieval as a set of conditions that determine which type of database should be copyrightable in the form of a component of contents-plus-indexes.

Copyrightable compilation is said to be of sufficient creativity, that is, originality in the form of a component of contents-plus-indexes. Its original way of categorization is represented in the selection of the type of index or metadata that is assigned to individual contents, or the type of taxonomy regarding indexes or metadata that integrate the individual contents into the entire content of each database.

The set of conditions on the original categorization regarding indexes or metadata is formulated as shown below: A categorization regarding indexes or metadata is original only when:

1. The type of index or metadata accepts discretionary selection in the domain of a problem database; otherwise,
2. The type of taxonomy regarding indexes or metadata accepts discretionary selection in the domain of a problem database.

We describe five typical cases of original or non-original categorization regarding indexes or metadata for keyword-based retrieval. Typical cases of non-original categorization include photo film cartridge database and white pages. Those cases do not accept any discretion in the selection of the type of index or metadata, or the type of taxonomy. The photo film cartridge database uses its respective film numbers as indexes for its retrieval operations. The taxonomy of the indexes is only based on page numbering. This kind of database does not have any other discretion in its selection of the type of index or taxonomy. The white pages extract metadata from telephone numbers and names of subscribers. The case of *Feist Publications, Inc. v. Rural Telephone Service* (1991) judged that its selection of the type of taxonomy was not discretionary because its alphabetical ordering of the metadata was a single alternative in that field of practice.

Meanwhile, the discretionary selection of the type of index or metadata, or taxonomy, constitutes copyrightable compilation of minimal creativity, that is, originality on the categorization regarding indexes or metadata. Typical cases of discretionary selection of the type of index or metadata include the Web document encyclopedia and the other types of telephone directory, such as yellow pages. The latter yellow pages satisfy minimal creativity as a copyrightable compilation. The white pages list telephone numbers of subscribers in alphabetical order, while the yellow pages list business phone numbers by a variety of categories and feature-classified advertisements of various sizes. In the case of discretionary selection of the type of taxonomy, even numerical indexing, such as page numbering, could form an original compilation, when page indexes work as categorization identifiers of a specific type of taxonomy in the domain of the problem database. For example, lawyers often identify individual cases by citing their page numbers without any reference to case titles or parties. Mere page numbers work as categorization identifiers as far as the selection of page numbers is based on a specific type of taxonomy and still allows other discretionary selection of the type of index, in this case, different ranges in paging or indexing of case numbers. The component of contents-plus-indexes, in particular,

cases-plus-pages, constitutes an original work of authorship as a copyrightable compilation as affirmed by the case of *West Publishing Co. v. Mead Data Central, Inc.* (1986/1987).

Let us suppose that a database restores pictures of starfish that are manually and numerically numbered by day/hour-chronicle interval based on their significant life stages from birth to death. That database is to be an original work of authorship as a copyrightable compilation in the form of a component of contents-plus-indexes, that is, pictures-plus-numbers.

We have formulated the set of conditions that determine which type of database is copyrightable as a multimedia digital library by assessing its original categorization regarding indexes or metadata whose collection identifies the entire content of the database as the referent of copyright protection. That condition set is effective in protecting the multimedia digital libraries that are associated with keyword-based retrieval approach for image retrieval.

PATENTABILITY OF CBIR MECHANISMS

CBIR realizes its retrieval operations by performing a number of "processes," that is, methods or means for data processing that constitute computer-related inventions in the form of computer programs. The computer-related inventions often combine means for data processing, some of which are prior disclosed inventions as computer programs. Meanwhile, the processes comprise the components for parameter setting that is adjusted to retrieve specific kinds of images in certain domains, particularly in the case of domain-specific approach of CBIR.

In CBIR, parametric values determine as thresholds which candidate image is similar to an exemplary requested image by computation of similarity of visual features (Rui et al., 1999; Smeulders et al., 2000; Yoshitaka & Ichikawa, 1999). A component for parameter setting realizes the thresholding operations in the form of a computer program with a set of ranges of parametric values. The parameter-setting component is familiar in mechanic inventions. Its typical example is a patented invention of an automatic temperature controller that adjusts body temperature of raw fish in cargo within certain range as fish are not frozen but kept chilled during the course of transportation. Inventors and practitioners demand a detailed set of the conditions for patenting the data-processing processes for CBIR in multimedia digital libraries.

PATENTING CBIR MECHANISMS

In this section, we technically interpret the scheme for patenting CBIR mechanisms in the form of the conditions on patentability. We focus on combinations of prior disclosed processes and parameter setting components. The formulated conditions consist of the following three requirements for patentability: "patentable subject matter" (entrance to patent protection), "non-obviousness" (technical advancement), and "enablement" (specification) (Merges, 1997).

For satisfying the requirement for patentable subject matter, the processes for performing CBIR functions must be claimed as the means for parameter setting, which perform certain retrieval functions. Otherwise, the discussed processes are considered not as specific inventions of the data-processing processes for performing CBIR functions but as the inventions that are peripheral or just related to general retrieval functions.

A data-processing process or method is patentable subject matter in the form of a computer-related invention, that is, a computer program (U.S. Patent Act, 2003; Jakes & Yoches, 1989) as far as the "specific machine …… produce(s) a useful, concrete, and tangible result … for transforming … " physical data ("*physical transformatio*n") (in *re Alappat*, 1994). A process must "perform

independent physical acts (post-computer process activities)," otherwise, "manipulate data representing physical objects or activities to achieve a practical application (pre-computer process activities)." CBIR operations automatically generate indexes as physical results on a computer and a memory, which require pre-and post-computer process activities as indispensable procedure through data processing between feature extraction and indexing, also between indexing and classification. Inventions should be of "technological arts." That requirement does not limit patentability of computer-related inventions because technological arts are equivalent to, in broad sense, the concept of useful or practical arts (Merges, 1997).

The requirement for non-obviousness on the combinations of the processes for data processing is listed as below:

1. The processes for performing CBIR functions must comprise the combinations of prior disclosed means to perform certain functions that are not predicated from any combination of the prior arts; in addition.
2. The processes for performing CBIR functions must realize quantitative and/or qualitative advancement.

Otherwise, the discussed processes are obvious so that they are not patentable as the processes for performing CBIR functions.

First, a combination of prior disclosed means should not be "suggested" from any disclosed means "with the reasonable expectation of success" (in *re Dow Chemical Co.*, 1988). Second, its asserted function must be superior to the conventional functions that are realized in the prior disclosed or patented means. On the latter issue, several solutions for performance evaluation are proposed and applicable. Müller et al. (2001) and Manchester Visualization Center (2000) proposed benchmarking of CBIR functions. Another general strategy is restriction of the scope of problem claims into a certain narrow field to which no prior arts have been applied. This claiming strategy is local optimization of application scope.

The requirement for enablement on the parameter setting components is listed as below:

1-a. The descriptions of the processes for performing CBIR functions must specify the formulas for parameter setting, otherwise,
1-b. The disclosed invention of the processes should have its co-pending application that describes the formulas in detail. In addition,
2-a. The processes must perform a new function as domain-general approach by a combination of the prior disclosed means; otherwise,
2-b. The processes as domain-specific approach should have improved formulas for parameter setting based on the prior disclosed means for performing CBIR functions and also examples of parametric values on parameter setting in descriptions.

For 2-b, the processes must specify the means for parameter setting by "giving a specific example of preparing an" application to enable those skilled in the arts to implement their best mode of the processes without undue experiment (*Autogiro Co. of America v. United States*, 1967; *Unique Concepts, Inc. v. Brown*, 1991). U.S. Patent and Trademark Office (1996a, 1996b) suggested that the processes comprising the means, that is, the components for parameter setting must disclose at least one of the following examples of parametric values on parameter setting:

1. Working or prophetic examples of initial values or weights on parameter setting.
2. Working examples of the ranges of parametric values on parameter setting.

The "working examples" are parametric values that are confirmed to work at actual laboratory or as prototype testing results. The "prophetic examples" are given without actual work by one skilled in the art.

It is a critical problem to define the scope of equivalent modification of process patents because parametric values in parameter setting are easy to modify and adjust at application. The scope of patent enforcement extends to what is equivalent to a claimed invention, as far as it is predictable from claims and descriptions (*Graver Tank & Mfg. Co. v. Linde Air Products Co.*, 1950; *Laitran Corp. v. Rexnord, Inc.*, 1991). Especially, domain-specific approach for performing CBIR functions must distinguish its claimed invention with its examples of parametric values from other improved formulas for parameter setting that are based on prior disclosed means. The scope of the equivalent modification of a patented process is defined within a certain specified scope as suggested from the exemplary parametric values.

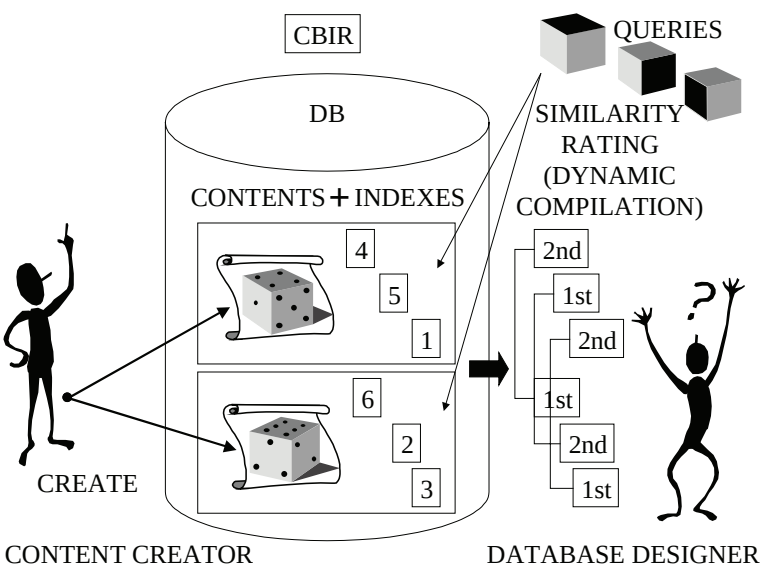
LIMITATIONS OF COPYRIGHT IN THE ADVENT OF CBIR

Two limitations on copyright are to be identified in its application to image databases as multimedia digital libraries:

1. A component of contents-plus-indexes does not identify the entire content of any database that is associated with CBIR operations; and,
2. Any copyrightable compilation in that database does not identify its entire content.

Figure 2 outlines the first limitation that the component of contents-plus-indexes, which is a collection of individual contents with *dynamic* indexes but does not identify the entire content of any database that is associated with CBIR operations. In the application of CBIR operations to a database, its component of contents-plus-indexes is just a collection of respectively rated and displayed individual contents whose similarity rating

Figure 2. A digital library of visual content associated with CBIR



order changes every time new sample images are requested as queries.

The second limitation is to be discovered in two phases. The discussed mere collection of individual contents in an image database is a static compilation without any minimal creativity in its categorization. That database is not copyrightable but so are its individual contents. Any proposed scheme does not offer the remedy for that problem at the present. The European Union legislated and executed a scheme for protecting databases, known as the *sui generis* right of database protection (Reinbothe, 1999; Samuelson, 1996). Its fundamental concept is based on the property of copyrightable compilation so that it does not protect the image databases with CBIR operations under the same reasons of the discussed limitation of the copyright protection. An only resort for a new scheme must be discovered in retrieval mechanisms that are to identify the entire content of each image database, multimedia digital library, as the referent of its protection.

PROMISING DIRECTION FOR PROTECTING MULTIMEDIA DIGITAL LIBRARIES WITH CBIR

Another new scheme must realize that the protection of multimedia digital libraries is to be leveraged by using the intellectual property of their retrieval mechanisms or data processing methodologies but not the copyright of their referents, that is, databases.

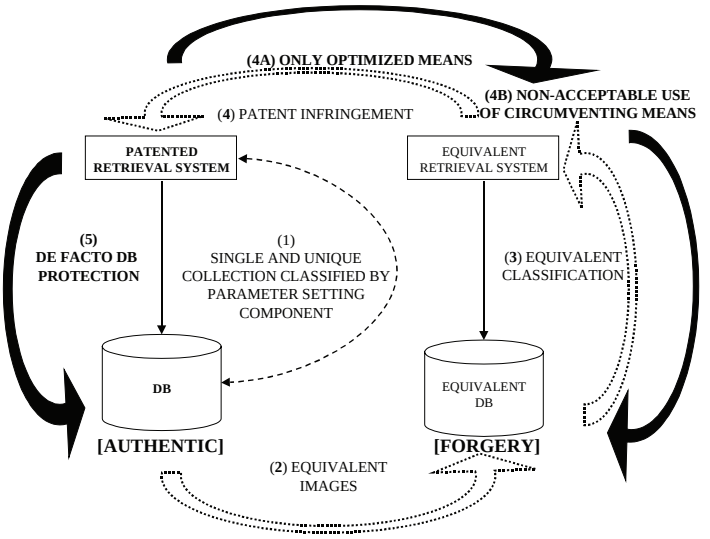
The practice in the field of bioinformatics offers a typical case of that kind of protection.

The methodologies for identifying genetic codes have been yielded with process patent in the form of computer programs, even before the admission of the patentability of genomic DNA sequences per se as material patent. A kind of exclusive protection over its referents, the gene per se, is leveraged by patenting the component of the data processing methodology that identifies a certain genetic code that is to represent the common and unique elements of genomic DNA sequences encoding a certain protein. Without such that patentable invention, any other means neither identifies the genetic code that corresponds to a specific gene nor represents the single and unique collection of genomic DNA sequences encoding a protein.

If certain conditions let a component in retrieval mechanisms identify a classification of the entire content of an image database as a single and unique collection in a specific domain, patenting that component realizes the same *de facto* protection of its referent, that is, the database with CBIR, as patenting the discussed component in bioinformatics does. It is based on a logic that CBIR is optimized in that specific domain of the image database as far as the patented component of the methodology integrates or identifies the classification of its referents (individual contents) into the single and unique collection (the entire content) (Sasaki & Kiyoki, 2003, 2002a).

That component must be discovered from consulting the technical properties of CBIR mechanisms, especially, domain-specific approach whose application focuses on the narrower or specific domains for image retrieval. In those domains, the feature extraction and index classification of the referent images are realized by optimizing the components for thresholding operations that evaluate structural similarity of extracted visual features of the images to identify a classification of the entire content of an image database. The component for thresholding operations consists of the means for selecting and/or adjusting parametric values. The parameter-setting component identifies a classification of the entire content of its referent image database by optimizing the parametric values on feature extraction and index classification in each specific domain (Sasaki & Kiyoki, 2003). That parameter setting component is to be a computer-related invention in the form of computer programs (U.S. Patent and Trademark Office, 1996a, 1996b; Sasaki &Kiyoki, 2002b, 2002c, 2005).

Figure 3. The patent-based protection scheme



A specific domain has a *sign* that is a set of visual features, for example, shape, color, texture, region, and etcetera that are extracted out of mutually similar images (Smeulders et al., 2000). Signs bridge the gaps between the sets of sample images requested as queries and the classes of mutually similar candidate images in each specific domain. When the pre-defined ranges of parametric values identify the signs and a classification of an image database as a single and unique collection, the patentable parameter-setting component leverages the de facto protection of the entire content of an image database in its referent specific domain. *Figure 3* outlines the scheme that leverages the de facto protection of an image database as a multimedia digital library in a certain specific domain:

1. Identify a specific narrow domain of an authentic image database, and assure its parameter setting component of the patented methodology to identify a classification of the entire content of an image database as a single and unique collection in a specific domain.
2. Assemble another forgery image database in the same or equivalent domain by duplicating, otherwise, restoring images that are similar or equivalent to the images of the authentic image database.
3. Discover that forgery image database to implement a circumventing methodology that classifies and identifies the same or equivalent classes of mutually similar images as the authentic patented methodology does.
4. Verify that the circumventing methodology is the equivalents to the patented methodology and infringes that methodology under the logic that without the methodology for identification no other method for retrieval is optimized in the discussed domain.
5. Realize the de facto protection of the authentic image database as is leveraged by the protection of the patented methodology for domain-specific approach of CBIR in the specific domain.

An advantage of the above patent-based protection is to use the registered right for leveraging

its de facto protection of its referent, database. Another merit is to restrict its exclusive protection leveraged by patent protection in the modest scope of specific domain. It is a well-known fact that parametric values are easy to modify and adjust in the application. Exemplary parametric values must be specified to define a clear boundary of the scope of equivalent modification of the claimed methodologies. As suggested from exemplary parametric values, the scope of modification is restricted within respective specific domains. The most significant problem in the protection of digital libraries is the preemption of application fields by the foregoing that are often the western countries. Reinbothe (1999) mentioned that any prospective protection of digital libraries should be fair to both the developers and the followers. The proposed scheme is not to have any excessive protection over a number of image databases as multimedia digital libraries in general domains.

CONCLUSION

In this chapter, we have discussed the legal issues on multimedia digital libraries and retrieval mechanisms, and technically interpreted the legal schemes for protecting multimedia digital libraries and retrieval mechanisms. The protection of its intellectual property right is a critical issue in the digital library community. We have presented the schemes for copyrighting the multimedia digital libraries with keyword-based retrieval, patenting the CBIR mechanisms, and protecting the multimedia digital libraries with CBIR.

REFERENCES

American Dental Association v. Delta Dental Plan Association, 126 F.3d 977 (7th Cir. 1997).

Autogiro Co. of America v. United States, 384 F.2d 391, 155 U.S.P.Q. 697 (Ct. Cl. 1967).

Borgman, C.L. (2000). *From Gutenberg to the global information infrastructure: Access to information in the networked worl*d. Digital Libraries and Electronic Publishing. Cambridge, MA: MIT Press.

Feist Publications, Inc. v. Rural Telephone Service, 499 U.S. 340, 111 S.Ct. 1282, 113 L.Ed.2d 358 (1991).

Gorman, R.A., & Ginsburg, J.C. (1993). *Copyright for the nineties: Cases and materials* (4th ed.). Contemporary legal education series. Charlottesville, NC: The Michie Company.

Graver Tank & Mfg. Co. v. Linde Air Products Co., 339 U.S. 605 (1950).

In *re Alappat*, 33 F.3d 1526, 31 U.S.P.Q.2d 1545 (Fed. Cir. 1994) (en banc).

In *re Dow Chemical Co.*, 837 F.2d 469, 473, 5 U.S.P.Q.2d 1529, 1531 (Fed. Cir. 1988).

Jakes, J.M., & Yoches, E.R. (1989). Legally speaking: Basic principles of patent protection for computer science. *Communications of the AC*M, *32* (8), 922–924.

Laitran Corp. v. Rexnord, Inc., 939 F.2d 1533, 19 U.S.P.Q.2d (BNA) 1367 (Fed. Cir. 1991).

Manchester Visualization Center. (2000). CBIR evaluation. Retrieved from the World Wide Web: *http://www.man.ac.uk/MVC/research/CBIR/*

Merges, R.P. (1997). *Patent law and policy: Cases and materials* (2nd ed.). Contemporary legal education series. Charlottesville, NC: The Michie Company.

Mûller, H., Mûller, W., Squire, D.M., March-and-Maillet, S., & Pun, T. (2001). Automated benchmarking in content-based image retrieval.

In *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo (ICME 2001*), (pp. 321–324). Tokyo, Japan.

Nimmer, M., Marcus, P., Myers, D., & Nimmer, D. (1991). *Cases and materials on copyright* (4th ed.). St. Paul, MN: West Publishing.

Reinbothe, J. (1999). The legal protection of non-creative databases. In *Proceedings of the Database Workshop of the International Conference of Electronic Commerce and Intellectual Property*, (WIPO/EC/CONF/99/SPK/22-A). Geneva, Switzerland, September 14-16. WIPO.

Rui, Y., Huang, T.S., & Chang, S.F. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representatio*n, *10* (4), 39–62.

Samuelson, P. (1996). Legally speaking: Legal protection for database content. *Communications of the ACM, 39* (12), 17–23.

Sasaki, H., & Kiyoki, Y. (2002a). Media kensaku enzin no tokkyo shutoku niyoru maruchimedia-detabesu no kenri hogo hoshiki [A methodology to protect a multi-media database by a patentable program of indexing and retrieval based-on semantic similarity]. *IPS of Japan Transactions on Databases, 43* (13), 108–127.

Sasaki, H., & Kiyoki, Y. (2002b). Patenting advanced search engines of multimedia databases. In S. Lesavich (ed.), *Proceedings of the 3rd International Conference on Law and Technology* (pp. 34–39). Cambridge, MA, November 6–7. International Society of Law and Technology (ISLAT). Anaheim, Calgary, Zurich: Acta Press.

Sasaki, H., & Kiyoki, Y. (2002c). Patenting the processes for content-based retrieval in digital libraries. In E.P. Lim, S. Foo, C. Khoo, H. Chen, E. Fox, S. Urs, & T. Costantino (eds.), *Proceedings of the 5th International Conference on Asian Digital Libraries (ICADL) – Digital Libraries: People, Knowledge, and Technology, Lecture Notes in Computer Science,* 2555 (pp. 471–482). Singapore, December 11–14. Berlin: Springer-Verlag.

Sasaki, H., & Kiyoki, Y. (2003). A proposal for digital library protection. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries* (p. 392). Houston, TX, May 27–31. Los Alamitos, CA: IEEE Computer Society Press.

Sasaki, H., & Kiyoki, Y. (2005). A formulation for patenting content-based retrieval processes in digital libraries. *Information Processing and Management, 41*(1), 57-74.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22* (12), 1349–1380.

Unique Concepts, Inc. v. Brown, 939 F.2d 1558, 19 U.S.P.Q.2d 1500 (Fed. Cir. 1991).

U.S. Copyright Act, 17 U.S.C. Sec. 101, & 103 (1976).

U.S. Patent Act, Title 35 U.S.C. Sec. 101, 103, & 112 (2003).

U.S. Patent and Trademark Office (1996a). *Examination guidelines for computer-related invention*s, 61 Fed. Reg. 7478 (Feb. 28, 1996) ("*Guideline*s"). Retrieved from the World Wide Web: *http://www.uspto.gov/web/offices/pac/dapp/oppd/patoc.htm*.

U.S. Patent and Trademark Office. (1996b). *Examination guidelines for computer-related inventions training materials directed to business, artificial intelligence, and mathematical processing applications ("Training Materials")*. Retrieved from the World Wide Web: *http://www.uspto.gov/web/offices/pac/compexam/examcomp.htm*.

West Publishing Co. v. Mead Data Central, Inc., 799 F.2d 1219 (8th Cir. 1986), cert. denied, 479 U.S. 1070 (1987).

Yoshitaka, A., & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering, 11* (1), 81–93.

This work was previously published in Design and Usability of Digital Libraries: Case Studies in the Asia Pacific, edited by Y.-C. Theng and S. Foo, pp. 238-253, copyright 2005 by Information Science Publishing (an imprint of IGI Global).

Chapter 1.24
Data Hiding in Document Images

Minya Chen
Polytechnic University, USA

Nasir Memon
Polytechnic University, USA

Edward K. Wong
Polytechnic University, USA

ABSTRACT

With the proliferation of digital media such as images, audio, and video, robust digital watermarking and data hiding techniques are needed for copyright protection, copy control, annotation, and authentication of document images. While many techniques have been proposed for digital color and grayscale images, not all of them can be directly applied to binary images in general and document images in particular. The difficulty lies in the fact that changing pixel values in a binary image could introduce irregularities that are very visually noticeable. Over the last few years, we have seen a growing but limited number of papers proposing new techniques and ideas for binary image watermarking and data hiding. In this chapter we present an overview and summary of recent developments on this important topic, and discuss important issues such as robustness and data hiding capacity of the different techniques.

INTRODUCTION

Given the increasing availability of cheap yet high quality scanners, digital cameras, digital copiers, printers and mass storage media the use of document images in practical applications is becoming more widespread. However, the same technology that allows for creation, storage and processing of documents in digital form, also provides means for mass copying and tampering of documents. Given the fact that digital documents need to be exchanged in printed format for many practical applications, any security mechanism for protecting digital documents would have to be compatible with the paper-based infrastructure. Consider for

example the problem of authentication. Clearly an authentication tag embedded in the document should survive the printing process. That means that the authentication tag should be embedded inside the document data rather than appended to the bitstream representing the document. The reason is that if the authentication tag is appended to the bitstream, a forger could easily scan the document, remove the tag, and make changes to the scanned copy and then print the modified document.

The process of embedding information into digital content without causing perceptual degradation is called *data hiding.* A special case of data hiding is *digital watermarking* where the embedded signal can depend on a secret key. One main difference between data hiding and watermarking is in whether an active adversary is present. In watermarking applications like copyright protection and authentication, there is an active adversary that would attempt to remove, invalidate or forge watermarks. In data hiding there is no such active adversary as there is no value associated with the act of removing the hidden information. Nevertheless, data hiding techniques need to be robust against accidental distortions.

A special case of data hiding is *steganography* (meaning *covered writing* in Greek), which is the science and art of secret communication. Although steganography has been studied as part of cryptography for many decades, the focus of steganography is secret communication. In fact, the modern formulation of the problem goes by the name of the *prisoner's problem*. Here Alice and Bob are trying to hatch an escape plan while in prison. The problem is that all communication between them is examined by a warden, Wendy, who will place both of them in solitary confinement at the first hint of any suspicious communication. Hence, Alice and Bob must trade seemingly inconspicuous messages that actually contain hidden messages involving the escape plan. There are two versions of the problem that are usually discussed — one where the warden is *passive*, and only observes messages, and the other where the warden is *active* and modifies messages in a limited manner to guard against hidden messages. The most important issue in steganography is that the very presence of a hidden message must be concealed. Such a requirement is not critical in general data hiding and watermarking problems.

Before we describe the different techniques that have been devised for data hiding, digital watermarking and steganography for document images, we briefly list different applications that would be enabled by such techniques.

1. *Ownership assertion.* To assert ownership of a document, Alice can generate a watermarking signal using a secret private key, and embed it into the original document. She can then make the watermarked document publicly available. Later, when Bob contends the ownership of a copy derived from Alice's original, Alice can produce the unmarked original and also demonstrate the presence of her watermark in Bob's copy. Since Alice's original is unavailable to Bob, he cannot do the same provided Alice has embedded her watermark in the proper manner (Holliman & Memon, 2000). For such a scheme to work, the watermark has to survive operations aimed at malicious removal. In addition, the watermark should be inserted in such a manner that it cannot be forged, as Alice would not want to be held accountable for a document that she does not own (Craver et al., 1998).
2. *Fingerprinting.* In applications where documents are to be electronically distributed over a network, the document owner would like to discourage unauthorized duplication and distribution by embedding a distinct watermark (or a fingerprint) in each copy of the data. If, at a later point in time, unauthorized copies of the document are

found, then the origin of the copy can be determined by retrieving the fingerprint. In this application the watermark needs to be invisible and must also be invulnerable to deliberate attempts to forge, remove or invalidate. The watermark should also be resistant to collusion. That is, a group of k users with the same document but containing different fingerprints should not be able to collude and invalidate any fingerprint or create a copy without any fingerprint.

3. *Copy prevention or control.* Watermarks can also be used for copy prevention and control. For example, every copy machine in an organization can include special software that looks for a watermark in documents that are copied. On finding a watermark the copier can refuse to create a copy of the document. In fact it is rumored that many modern currencies contain digital watermarks which when detected by a compliant copier will disallow copying of the currency. The watermark can also be used to control the number of copy generations permitted. For example a copier can insert a watermark in every copy it makes and then it would not allow further copying when presented a document that already contains a watermark.
4. *Authentication.* Given the increasing availability of cheap yet high quality scanners, digital cameras, digital copiers and printers, the authenticity of documents has become difficult to ascertain. Especially troubling is the threat that is posed to conventional and well established document based mechanisms for identity authentication, like passports, birth certificates, immigration papers, driver's license and picture IDs. It is becoming increasingly easier for individuals or groups that engage in criminal or terrorist activities to forge documents using off-the-shelf equipment and limited resources. Hence it is important to ensure that a given document was originated from a specific source and that it has not been changed, manipulated or falsified. This can be achieved by embedding a watermark in the document. Subsequently, when the document is checked, the watermark is extracted using a unique key associated with the source, and the integrity of the data is verified through the integrity of the extracted watermark. The watermark can also include information from the original document that can aid in undoing any modification and recovering the original. Clearly a watermark used for authentication purposes should not affect the quality of the document and should be resistant to forgeries. Robustness is not critical, as removal of the watermark renders the content inauthentic and hence is of no value.
5. *Metadata Binding.* Metadata information embedded in an image can serve many purposes. For example, a business can embed the Web site URL for a specific product in a picture that shows an advertisement for that product. The user holds the magazine photo in front of a low-cost CMOS camera that is integrated into a personal computer, cellular phone, or a personal digital assistant. The data are extracted from the low-quality picture and is used to take the browser to the designated Web site. For example, in the mediabridge application (*http://www.digimarc.com*), the information embedded in the document image needs to be extracted despite distortions incurred in the print and scan process. However, these distortions are just a part of a process and not caused by an active and malicious adversary.

The above list represents example applications where data hiding and digital watermarks could potentially be of use. In addition, there are many other applications in digital rights management (DRM) and protection that can benefit from data

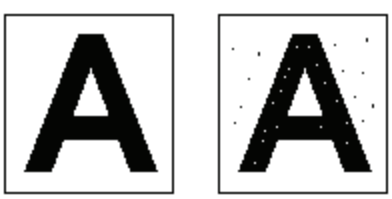
hiding and watermarking technology. Examples include tracking the use of documents, automatic billing for viewing documents, and so forth. From the variety of potential applications exemplified above it is clear that a digital watermarking technique needs to satisfy a number of requirements. Since the specific requirements vary with the application, data hiding and watermarking techniques need to be designed within the context of the entire system in which they are to be employed. Each application imposes different requirements and would require different types of watermarking schemes.

Over the last few years, a variety of digital watermarking and data hiding techniques have been proposed for such purposes. However, most of the methods developed today are for grayscale and color images (Swanson et al., 1998), where the gray level or color value of a selected group of pixels is changed by a small amount without causing visually noticeable artifacts. These techniques cannot be directly applied to binary document images where the pixels have either a 0 or a 1 value. Arbitrarily changing pixels on a binary image causes very noticeable artifacts (see Figure 1 for an example). A different class of embedding techniques must therefore be developed. These would have important applications in a wide variety of document images that are represented as binary foreground and background; for example, bank checks, financial instruments, legal documents, driver licenses, birth certificates, digital books, engineering maps, architectural drawings, road maps, and so forth. Until recently, there has been little work on watermarking and data hiding techniques for binary document images. In the remaining portion of this chapter we describe some general principles and techniques for document image watermarking and data hiding. Our aim is to give the reader a better understanding of the basic principles, inherent trade-offs, strengths, and weaknesses of document image watermarking and data hiding techniques that have been developed in recent years.

Most document images are binary in nature and consist of a foreground and a background color. The foreground could be printed characters of different fonts and sizes in text documents, handwritten letters and numbers in a bank check, or lines and symbols in engineering and architectural drawings. Some documents have multiple gray levels or colors, but the number of gray levels and colors is usually few and each local region usually has a uniform gray level or color, as opposed to the different gray levels and colors you find at individual pixels of a continuous-tone image. Some binary documents also contain grayscale images represented as half-tone images, for example the photos in a newspaper. In such images, *nxn* binary patterns are used to approximate gray level values of a gray scale image, where *n* typically ranges from two to four. The human visual system performs spatial integration of the fine binary patterns within local regions and perceives them as different intensities (Foley et al., 1990).

Many applications require that the information embedded in a document be recovered despite accidental or malicious distortions they

Figure 1. Effect of arbitrarily changing pixel values on a binary image



may undergo. Robustness to printing, scanning, photocopying, and facsimile transmission is an important consideration when hardcopy distributions of documents are involved. There are many applications where robust extraction of the embedded data is not required. Such embedding techniques are called *fragile* embedding techniques. For example, fragile embedding is used for authentication whereby any modification made to the document can be detected due to a change in the watermark itself or a change in the relationship between the content and the watermark. Fragile embedding techniques could also be used for steganography applications.

In the second section, of this chapter, we summarize recent developments in binary document image watermarking and data hiding techniques. In the third section, we present a discussion on these techniques, and in the fourth section we give our concluding remarks.

DATA HIDING TECHNIQUES FOR DOCUMENTS IMAGES

Watermarking and data hiding techniques for binary document images can be classified according to one of the following embedding methods: text line, word, or character shifting, fixed partitioning of the image into blocks, boundary modifications, modification of character features, modification of run-length patterns, and modifications of half-tone images. In the rest of this section we describe representative techniques for each of these methods.

Text Line, Word or Character Shifting

One class of robust embedding methods shifts a text line, a group of words, or a group of characters by a small amount to embed data. They are applicable to documents with formatted text.

S. Low and co-authors have published a series of papers on document watermarking based on line and word shifting (Low et al., 1995a, 1995b, 1998; Low & Maxemchuk, 1998; Maxemchuk & Low, 1997). These methods are applicable to documents that contain paragraphs of printed text. Data is embedded in text documents by shifting lines and words spacing by a small amount (1/150 inch.) For instance, a text line can be moved up to encode a '1' or down to encode a '0,' a word can be moved left to encode a '1' or right to encode '0'. The techniques are robust to printing, photocopying, and scanning. In the decoding process, distortions and noise introduced by printing, photocopying and scanning are corrected and removed as much as possible. Detection is by use of maximum-likelihood detectors. In the system they implemented, line shifts are detected by the change in the distance of the marked line and two control lines — the lines immediately above and below the marked line. In computing the distance between two lines, the estimated centroids of the horizontal profiles (projections) of the two lines are used as reference points. Vertical profiles (projections) of words are used for detecting word shifts. The block of words to be marked (shifted) is situated between two control blocks of words. Shifting is detected by computing the correlation between the received profile and the uncorrupted marked profile. The line shifting approach has low embedding capacity but the embedded data are robust to severe distortions introduced by processes such as printing, photocopying, scanning, and facsimile transmission. The word shifting approach has better data embedding capacity but reduced robustness to printing, photocopying and scanning.

In Liu et al. (1999), a combined approach that marks a text document by line or word shifting, and detects the watermark in the frequency domain by Cox et al.'s algorithm (Cox et al., 1996) was proposed. It attempts to combine the unobtrusiveness of spatial domain techniques and the good detection performance of frequency domain techniques. Marking is performed according to the line and word shifting method described above.

The frequency watermark X is then computed as the largest N values of the absolute differences in the transforms of the original document and the marked document. In the detection process, the transform of the corrupted document is first computed. The corrupted frequency watermark X^* is then computed as the largest N values of the absolute differences in the transform of the corrupted document and the original document. The detection of watermark is by computing a similarity between X and X^*. This method assumes that the transform of the original document, and the frequency watermark X computed from the original document and the marked document (before corruption) is available during the detection process.

In Brassil and O'Gorman (1996), it is shown that the height of a bounding box enclosing a group of words can be used to embed data. The height of the bounding box is increased by either shifting certain words or characters upward, or by adding pixels to end lines of characters with ascenders or descenders. The method was proposed to increase the data embedding capacity over the line and/or word shifting methods described above. Experimental results show that bounding box expansions as small as 1/300 inch can be reliably detected after several iterations of photocopying. For each mark, one or more adjacent words on an encodable text line are selected for displacement according to a selection criterion. The words immediately before and after the shifted word(s), and a block of words on the text line immediately above or below the shifted word(s), remain unchanged and are used as "reference heights" in the decoding process. The box height is measured by computing a local horizontal projection profile for the bounding box. This method is very sensitive to baseline skewing. A small rotation of the text page can cause distortions in bounding box height, even after de-skewing corrections. Proper methods to deal with skewing require further research.

In Chotikakamthorn (1999), character spacing is used as the basic mechanism to hide data. A line of text is first divided into blocks of characters. A data bit is then embedded by adjusting the widths of the spaces between the characters within a block, according to a predefined rule. This method has advantage over the word spacing method above in that it can be applied to written languages that do not have spaces with sufficiently large width for word boundaries; for example, Chinese, Japanese, and Thai. The method has embedding capacity comparable to that of the word shifting method. Embedded data are detected by matching character spacing patterns corresponding to data bits '0' or '1'. Experiments show that the method can withstand document duplications. However, improvement is needed for the method to be robust against severe document degradations. This could be done by increasing the block size for embedding data bits, but this also decreases the data embedding capacity.

Fixed Partitioning of Images

One class of embedding methods partitions an image into fixed blocks of size m x n, and computes some pixel statistics or invariants from the blocks for embedding data. They can be applied to binary document images in general; for example, documents with formatted text or engineering drawings.

In Wu et al. (2000), the input binary image is divided into 3x3 (or larger) blocks. The flipping priorities of pixels in a 3x3 block are then computed and those with the lowest scores can be changed to embed data. The flipping priority of a pixel is indicative of the estimated visual distortion that would be caused by flipping the value of a pixel from 0 to 1 or from 1 to 0. It is computed by considering the change in smoothness and connectivity in a 3x3 window centered at the pixel. Smoothness is measured by the horizontal, vertical, and diagonal transitions, and connectivity is measured by the number of black and white clusters in the

3x3 window. Data are embedded in a block by modifying the total number of black pixels to be either odd or even, representing data bits 1 and 0, respectively. Shuffling is used to equalize the uneven embedding capacity over the image. It is done by random permutation of all pixels in the image after identifying the flippable pixels.

In Koch and Zhao (1995), an input binary image is divided into blocks of 8x8 pixels. The numbers of black and white pixels in each block are then altered to embed data bits 1 and 0. A data bit 1 is embedded if the percentage of white pixels is greater than a given threshold, and a data bit 0 is embedded if the percentage of white pixels is less than another threshold. A group of contiguous or distributed blocks is modified by switching white pixels to black or vice versa until such thresholds are reached. For ordinary binary images, modifications are carried out at the boundary of black and white pixels, by reversing the bits that have the most neighbors with the opposite pixel value. For dithered images, modifications are distributed throughout the whole block by reversing bits that have the most neighbors with the same pixel value. This method has some robustness against noise if the difference between the thresholds for data bits 1 and 0 is sufficiently large, but this also decreases the quality of the marked document.

In Pan et al. (2000), a data hiding scheme using a secret key matrix K and a weight matrix W is used to protect the hidden data in a host binary image. A host image F is first divided into blocks of size mxn. For each block F_i, data bits $b_1 b_2 \ldots b_r$ are embedded by ensuring the invariant

$$SUM((F_i \oplus K) \otimes W) \equiv b_1 b_2 \ldots b_r \pmod{2^r},$$

where $\oplus$ represents the bit-wise exclusive OR operation, $\otimes$ represents pair-wise multiplication, and is the sum of all elements in a matrix. Embedded data can be easily extracted by computing:

$$SUM((F_i \oplus K) \otimes W) \pmod{2^r}$$

The scheme can hide as many as bits of data in each image block by changing at most two bits in the image block. It provides high security, as long as the block size (m x n) is reasonably large. In a 256x256 test image divided into blocks of size 4x4, 16,384 bits of information were embedded. This method does not provide any measure to ensure good visual quality in the marked document.

In Tseng and Pan (2000), an enhancement was made to the method proposed in Pan et al. (2000) by imposing the constraint that every bit that is to be modified in a block is adjacent to another bit that has the opposite value. This improves the visual quality of the marked image by making the inserted bits less visible, at the expense of sacrificing some data hiding capacity. The new scheme can hide up to bits of data in an mxn image by changing at most two bits in the image block.

Boundary Modifications

In Mei et al. (2001), the data are embedded in the eight-connected boundary of a character. A fixed set of pairs of five-pixel long boundary patterns were used for embedding data. One of the patterns in a pair requires deletion of the center foreground pixel, whereas the other requires the addition of a foreground pixel. A unique property of the proposed method is that the two patterns in each pair are dual of each other — changing the pixel value of one pattern at the center position would result in the other. This property allows easy detection of the embedded data without referring to the original document, and without using any special enforcing techniques for detecting embedded data. Experimental results showed that the method is capable of embedding about 5.69 bits of data per character (or connected component) in a full page of text digitized at 300 dpi. The method can be applied to general document images with connected components; for example, text documents or engineering drawings.

Modifications of Character Features

This class of techniques extracts local features from text characters. Alterations are then made to the character features to embed data.

In Amamo and Misaki (1999), text areas in an image are identified first by connected component analysis, and are grouped according to spatial closeness. Each group has a bounding box that is divided into four partitions. The four partitions are divided into two sets. The average width of the horizontal strokes of characters is computed as feature. To compute average stroke width, vertical black runs with lengths less than a threshold are selected and averaged. Two operations — "make fat" and "make thin" — are defined by increasing and decreasing the lengths of the selected runs, respectively. To embed a "1" bit, the "make fat" operation is applied to partitions belonging to set 1, and the "make thin" operation is applied to partitions belongs to set 2. The opposite operations are used to embed "0" bit. In the detection process, detection of text line bounding boxes, partitioning, and grouping are performed. The stroke width features are extracted from the partitions, and added up for each set. If the difference of the sum totals is larger than a positive threshold, the detection process outputs 1. If the difference is less than a negative threshold, it outputs 0. This method could survive the distortions caused by print-and-scan (re-digitization) processes. The method's robustness to photocopying needs to be furthered investigated.

In Bhattacharjya and Ancin (1999), a scheme is presented to embed secret messages in the scanned grayscale image of a document. Small sub-character-sized regions that consist of pixels that meet criteria of text-character parts are identified first, and the lightness of these regions are modulated to embed data. The method employs two scans of the document — a low resolution scan and a high resolution scan. The low-resolution scan is used to identify the various components of the document and establish a coordinate system based on the paragraphs, lines and words found in the document. A list of sites for embedding data is selected from the low resolution scanned image. Two site selection methods were presented in the paper. In the first method, a text paragraph is partitioned into grids of 3x3 pixels. Grid cells that contain predominately text-type pixels are selected. In the second method, characters with long strokes are identified. Sites are selected at locations along the stroke. The second scan is a full-resolution scan that is used to generate the document copy. The pixels from the site lists generated in the low-resolution scan are identified and modulated by the data bits to be embedded. Two or more candidate sites are required for embedding each bit. For example, if the difference between the average luminance of the pixels belonging to the current site and the next one is positive, the bit is a 1; else, the bit is a 0. For robustness, the data to be embedded are first coded using an error correcting code. The resulting bits are then scrambled and dispersed uniformly across the document page. For data retrieval, the average luminance for the pixels in each site is computed and the data are retrieved according to the embedding scheme and the input site list. This method was claimed to be robust against printing and scanning. However, this method requires that the scanned grayscale image of a document be available. The data hiding capacity of this method depends on the number of sites available on the image, and in some cases, there might not be enough sites available to embed large messages.

Modification of Run-Length

In Matsui and Tanaka (1994), a method was proposed to embed data in the run-lengths of facsimile images. A facsimile document contains 1,728 pixels in each horizontal scan line. Each run length of black (or foreground) pixels is coded using modified Huffman coding scheme according to the statistical distribution of run-lengths. In the proposed method, each run length of black pixels

is shortened or lengthened by one pixel according to a sequence of signature bits. The signature bits are embedded at the boundary of the run lengths according to some pre-defined rules.

Modifications of Half-Toned Images

Several watermarking techniques have been developed for half-tone images that can be found routinely in printed matters such as books, magazines, newspapers, printer outputs, and so forth. This class of methods can only be used for half-tone images, and are not suitable for other types of document images. The methods described in Baharav and Shaked (1999) and Wang (2001) embed data during the half-toning process. This requires the original grayscale image. The methods described in Koch and Zhao (1995) and Fu and Au (2000a, 2000b, 2001) embed data directly into the half-tone images after they have been generated. The original grayscale image is therefore not required.

In Baharav and Shaked (1999), a sequence of two different dither matrices (instead of one) was used in the half-toning process to encode the watermark information. The order in which the two matrices are applied is the binary representation of the watermark. In Knox (United States Patent) and Wang (United States Patent), two screens were used to form two halftone images and data were embedded through the correlations between the two screens.

In Fu and Au (2000a, 2000b), three methods were proposed to embedded data at pseudo-random locations in half-tone images without knowledge of the original multi-tone image and the half-toning method. The three methods, named DHST, DHPT, and DHSPT, use one half-tone pixel to store one data bit. In DHST, *N* data bits are hidden at *N* pseudo-random locations by forced toggling. That is, when the original half-tone pixel at the pseudo-random locations differs from the desired value, it is forced to toggle. This method results in undesirable clusters of white or black pixels. In the detection process, the data are simply read from the *N* pseudo-random locations. In DHPT, a pair of white and black pixels (instead of one in DHST) is chosen to toggle at the pseudo-random locations. This improves over DHST by preserving local intensity and reducing the number of undesirable clusters of white or black pixels. DHSPT improves upon DHPT by choosing pairs of white and black pixels that are maximally connected with neighboring pixels before toggling. The chosen maximally connected pixels will become least connected after toggling and the resulting clusters will be smaller, thus improving visual quality.

In Fu and Au (2001), an algorithm called *intensity selection* (IS) is proposed to select the best location, out of a set of candidate locations, for the application of the DHST, DHPT and DHSPT algorithms. By doing so, significant improvement in visual quality can be obtained in the output images without sacrificing data hiding capacity. In general, the algorithm chooses pixel locations that are either very bright or very dark. It represents a data bit as the parity of the sum of the half-tone pixels at *M* pseudo-random locations and selects the best out of the *M* possible locations. This algorithm, however, requires the original grayscale image or computation of the inverse-half-toned image.

In Wang (2001), two data hiding techniques for digital half-tone images were described: *modified ordered dithering* and *modified multiscale error diffusion*. In the first method, one of the 16 neighboring pixels used in the dithering process is replaced in an ordered or pre-programmed manner. The method was claimed to be similar to replacing the insignificant one or two bits of a grayscale image, and is capable of embedding 4,096 bits in an image of size 256 x 256 pixels. The second method is a modification of the *multi-scale error diffusion* (MSED) algorithm for half-toning as proposed in Katsavounidis and Kuo (97), which alters the binarization sequence of the error diffusion process based on the global and

local properties of intensity in the input image. The modified algorithm uses fewer floors (e.g., three or four) in the image pyramid and displays the binarization sequence in a more uniform and progressive way. After 50% of binarization is completed, the other 50% is used for encoding the hidden data. It is feasible that edge information can be retained with this method.

Kacker and Allebach propose a joint halftoning and watermarking approach (Kacker & Allebach, 2003), that combines optimization based halftoning with a spread spectrum robust watermark. The method uses a joint metric to account for the distortion between a continuous tone and a halftone (FWMSE), as well as a watermark detectability criterion (correlation). The direct binary search method (Allebach et al., 1994) is used for searching a halftone that minimizes the metric. This method is obviously extendable in that other distortion metric and/or watermarking algorithms can be used.

DISCUSSION

Robustness to printing, scanning, photocopying, and facsimile transmission is an important consideration when hardcopy distributions of documents are involved. Of the methods described above, the line and word shifting approaches described in Low et al. (1995a, 1995b, 1998), Maxemchuk and Low (1997), Low and Maxemchuk (1998), and Liu et al. (1999), and the method using intensity modulation of character parts (Bhattacharjya & Ancin, 1999) are reportedly robust to printing, scanning, and photocopying operations. These methods, however, have low data capacity. The method described in Amamo and Misaki (1999) reportedly can survive printing and scanning (re-digitization) if the strokes remain in the image. This method's robustness to photocopying still needs to be determined. The bounding box expansion method described in Brassil and O'Gorman (1996) is a robust technique, but further research is needed to develop an appropriate document deskewing technique for the method to be useful. The character spacing width sequence coding method described in Chotikakamthorn (1999) can withstand a modest amount of document duplications.

The methods described in Wu et al. (2000), Pan et al. (2000), Tseng and Pan (2000), Mei et al. (2001), Matsui and Tanaka(1994), Wang (2001), and Fu and Au (2000a, 200b, 2001) are not robust to printing, scanning and copying operations but they offer high data embedding capacity. These methods are useful in applications when documents are distributed in electronic form, when no printing, photocopying, and scanning of hardcopies are involved. The method in Koch and Zhao (1995) also has high embedding capacity. It offers some amount of robustness if the two thresholds are chosen sufficiently apart, but this also decreases image quality.

Methods based on character feature modifications require reliable extraction of the features. For example, the methods described in Amamo and Misaki (1999) and one of the two site-selection methods presented in Bhattacharjya and Anci (1999) require reliable extraction of character strokes. The boundary modification method presented in Mei et al. (2001) traces the boundary of a character (or connected-component), which can always be reliably extracted in binary images. This method also provides direct and good image quality control. The method described in Matsui and Tanaka (1994) was originally developed for facsimile images, but could be applied to regular binary document images. The resulting image quality, however, may be reduced.

A comparison of the above methods shows that there is a trade off between embedding capacity and robustness. Data embedding capacity tends to decrease with increased robustness. We also observed that for a method to be robust, data must be embedded based on computing some statistics

over a reasonably large set of pixels, preferably spread out over a large region, instead of based on the exact locations of some specific pixels. For example, in the line shifting method, data are embedded by computing centroid position from a horizontal line of text pixels, whereas in the boundary modification method, data are embedded based on specific configurations of a few boundary pixel patterns.

In addition to robustness and capacity, another important characteristic of a data hiding technique is its "security" from a steganographic point of view. That is, whether documents that contain an embedded message can be distinguished from documents that do not contain any message. Unfortunately, this aspect has not been investigated in the literature. However, for any of the above techniques to be useful in a covert communication application, the ability of a technique to be indistinguishable is quite critical. For example, a marked document created using line and word shifting can easily be spotted as it has characteristics that are not expected to be found in "normal" documents. The block-based techniques and boundary-based technique presented in the second section may produce marked documents that are distinguishable if they introduce too many irregularities or artifacts. This needs to be further investigated. A similar comment applies to the techniques presented in the second section. In general, it appears that the development of "secure" steganography techniques for binary documents has not received enough attention in the research community and much work remains to be done in this area.

Table 1 summarizes the different methods in terms of embedding techniques, robustness,

Table 1. Comparison of techniques

Techniques	**Robustness**	**Advantages (+) / Disadvantages (-)**	**Capacity**	**Limitations**
Line shifting	High		Low	Formatted text only
Word shifting	Medium		Low/Medium	Formatted text only
Bounding box expansion	Medium -	Sensitive to document skewing	Low/Medium	Formatted text only
Character spacing	Medium +	Can be applied to languages with no clear-cut word boundaries	Low/Medium	Formatted text only
Fixed partitioning -- Odd/Even pixels	None +	Can be applied to binary images in general	High	
Fixed partitioning -- Percentage of white/black pixels	Low/Medium +	Can be applied to binary images in general - Image quality may be reduced	High	
Fixed partitioning -- Logical invariant	None +	Embed multiple bits within each block + Use of a secret key	High	
Boundary modifications	None +	Can be applied to general binary images + Direct control on image quality	High	

advantages/disadvantages, data embedding capacity, and limitations. Robustness here refers to robustness to printing, photocopying, scanning, and facsimile transmission.

CONCLUSION

We have presented an overview and summary of recent developments in binary document image watermarking and data hiding research. Although there has been little work done on this topic until recent years, we are seeing a growing number of papers proposing a variety of new techniques and ideas. Research on binary document watermarking and data hiding is still not as mature as for color and grayscale images. More effort is needed to address this important topic. Future research should aim at finding methods that offer robustness to printing, scanning, and copying, yet provide good data embedding capacity. Quantitative methods should also be developed to evaluate the quality of marked images. The steganographic capability of different techniques needs to be investigated and techniques that can be used in covert communication applications need to be developed.

REFERENCES

Allebach, J.P., Flohr, T.J., Hilgenberg, D.P., & Atkins, C.B. (1994, May). Model-based halftoning via direct binary search. *Proceedings of IS&T's 47th Annual Conference,* (pp. 476-482), Rochester, NY.

Amamo, T., & Misaki, D. (1999). Feature calibration method for watermarking of document images. *Proceedings of 5th Int'l Conf on Document Analysis and Recognition,* (pp. 91-94), Bangalore, India.

Baharav, Z., & Shaked, D. (1999, January). Watermarking of dither half-toned images. *Proc. of SPIE Security and Watermarking of Multimedia Contents, 1,*307-313.

Bhattacharjya, A.K., & Ancin, H. (1999). Data embedding in text for a copier system. *Proceedings of IEEE International Conference on Image Processing, 2,* 245-249.

Brassil, J., & O'Gorman, L. (1996, May). Watermarking document images with bounding box expansion. *Proceedings of 1st Int'l Workshop on Information Hiding,* (pp. 227-235). Newton Institute, Cambridge, UK.

Chotikakamthorn, N. (1999). Document image data hiding techniques using character spacing width sequence coding. *Proc. IEEE Intl. Conf. Image Processing,* Japan.

Cox, I., Kilian, J., Leighton, T., & Shamoon, T. (1996, May/June). Secure spread spectrum watermarking for multimedia. In R. Anderson (Ed.), *Proc. First Int. Workshop Information Hiding* (pp. 183-206). Cambridge, UK: Springer-Verlag.

Craver, S., Memon, N., Yeo, B., & Yeung, M. (1998, May). Resolving rightful ownership with invisible watermarking techniques: Limitations, attacks, and implications. *IEEE Journal on Selected Areas in Communications, 16*(4), 573-586.

Digimarc Corporation. http://www.digimarc.com.

Foley, J.D., Van Dam, A., Feiner, S.K., & Hughes, J.F. (1990). *Computer graphics: Principles and practice* (2nd ed.). Addison-Wesley.

Fu, M.S., & Au, O.C. (2000a, January). Data hiding for halftone images. *Proc of SPIE Conf. On Security and Watermarking of Multimedia Contents II, 3971,* 228-236.

Fu, M.S., & Au, O.C. (2000b, June 5-9). Data hiding by smart pair toggling for halftone images. *Proc. of IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, 4,* (pp. 2318-2321).

Fu, M.S., & Au, O.C. (2001). Improved halftone image data hiding with intensity selection. *Proc. IEEE International Symposium on Circuits and Systems, 5,* 243-246.

Holliman, M., & Memon, N. (2000, March). Counterfeiting attacks and blockwise independent watermarking techniques. *IEEE Transactions on Image Processing, 9*(3), 432-441.

Kacker, D., & Allebach, J.P. (2003, April). Joint halftoning and watermarking. *IEEE Trans. Signal Processing, 51,* 1054-1068.

Katsavounidis, I., & Jay Kuo, C.C. (1997, March). A multiscale error diffusion technique for digital half–toning. *IEEE Trans. on Image Processing, 6*(3), 483-490.

Knox, K.T. *Digital watermarking using stochastic screen patterns,* United States Patent Number 5,734,752.

Koch, E., & Zhao, J. (1995, August). Embedding robust labels into images for copyright protection. *Proc. International Congress on Intellectual Property Rights for Specialized Information, Knowledge & New Technologies,* Vienna.

Liu, Y., Mant, J., Wong, E., & Low, S.H. (1999, January).Marking and detection of text documents using transform-domain techniques. *Proc. SPIE Conf. on Security and Watermarking of Multimedia Contents,* (pp. 317-328), San Jose, CA.

Low, S.H., Lapone, A.M., & Maxmchuk, N.F. (1995, November 13-17). Document identification to discourage illicit copying. *IEEE GlobeCom 95,* Singapore.

Low, S.H., & Maxemchuk, N.F. (1998, May). Performance comparison of two text marking methods. *IEEE Journal on Selected Areas in Communications, 16*(4).

Low, S.H., Maxemchuk, N.F., Brassil, J.T., & O'Gorman, L. (1995). Document marking and identification using both line and word shifting. *Infocom 95.* Los Alamitos, CA: IEEE Computer Society Press.

Low, S.H., Maxemchuk, N.F., & Lapone, A.M. (1998, March). Document identification for copyright protection using centroid detection. *IEEE Trans. on Comm., 46*(3), 372-83.

Matsui, K. & Tanaka, K. (1994). Video-steganography: How to secretly embed a signature in a picture. *Proceedings of IMA Intellectual Property Project, 1*(1), 187-206.

Maxemchuk, N.F., & Low, S.H. (1997, October). Marking text documents. *Proceedings of IEEE Intl Conference on Image Processing.*

Mei, Q., Wong, E.K., & Memon, N. (2001, January). Data hiding in binary text documents. *SPIE Proc Security and Watermarking of Multimedia Contents III,* San Jose, CA.

Pan, H.-K., Chen, Y.-Y., & Tseng, Y.-C. (2000). A secure data hiding scheme for two-color images. *IEEE Symposium on Computers and Communications.*

Swanson, M., Kobayashi, M., & Tewfik, A. (1998, June). Multimedia data embedding and watermarking technologies. *IEEE Proceedings, 86*(6), 1064-1087.

Tseng, Y., & Pan, H. (2000). Secure and invisible data hiding in 2-color images. *IEEE Symposium on Computers and Communications.*

Wang, H.-C.A. (2001, April 2-4). Data hiding techniques for printed binary images. *The International Conference on Information Technology: Coding and Computing.*

Wang, S.G. *Digital watermarking using conjugate halftone screens,* United States Patent Number 5,790,703.

Wu, M., Tang, E., & Liu, B. (2000, July 31-August 2). Data hiding in digital binary images. *Proc. IEEE Int'l Conf. on Multimedia and Expo,* New York.

This work was previously published in Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property, edited by C.-S. Lu, pp.231-247, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.25
Emergent Semantics: An Overview

Viranga Ratnaike
Monash University, Australia

Bala Srinivasan
Monash University, Australia

Surya Nepal
CSIRO ICT Centre, Australia

ABSTRACT

The semantic gap is recognized as one of the major problems in managing multimedia semantics. It is the gap between sensory data and semantic models. Often the sensory data and associated context compose situations which have not been anticipated by system architects. Emergence is a phenomenon that can be employed to deal with such unanticipated situations. In the past, researchers and practitioners paid little attention to applying the concepts of emergence to multimedia information retrieval. Recently, there have been attempts to use emergent semantics as a way of dealing with the semantic gap. This chapter aims to provide an overview of the field as it applies to multimedia. We begin with the concepts behind emergence, cover the requirements of emergent systems, and survey the existing body of research.

INTRODUCTION

Managing media semantics should not necessarily involve semantic descriptions or classifications of media objects for future use. Information needs, for a user, can be task dependent, with the task itself evolving and not known beforehand. In such situations, the semantics and structure will also evolve, as the user interacts with the content, based on an abstract notion of the information required for the task. That is, users can interpret multimedia content, in context, at the time of information need. One way to achieve this is through a field of study known as *emergent semantics.*

Emergence is the phenomenon of complex structures arising from interactions between simple units. Properties or features appear that were not previously observed as functional characteristics of the units. Though constraints on a system can influence the formation of the emergent structure, they do not directly describe it. While emergence is a new concept in multimedia, it has been used in fields such as biology, physics and economics, as well as having a rich philosophical history. To the best of our knowledge, commercial emergent systems do not currently exist. However, there is research into the various technologies that would be required. This chapter aims to outline some characteristics of emergent systems and relevant tools and techniques.

The foundation for computational emergence is found in the constrained generating procedures (CGP) of John Holland (Holland, 2000). Initially there are only simple units and mechanisms. These mechanisms interact to form complex mechanisms, which in turn interact to form very complex mechanisms. This interaction results in self-organization through synthesis. If we relate the concept of CGP to multimedia, the simple units are sensory data, extracted features or even multimedia objects. Participating units can also come from other sources such as knowledge bases. Semantic emergence occurs when meaningful behaviour (phenotype) or complex semantic representation (genotype) arises from the interaction of these units. This includes user interaction, the influence of context, and relationships between media.

Context helps to deal with the problem of subjectivity, which occurs when there are multiple interpretations of a multimedia instance. World knowledge and context help to select one interpretation from the many. Ideally, we want to form semantic structures that can be understood by third parties who do not have access to the multimedia instance. This is not the same as relevance to the user. A system might want to determine what is of interest to one user, and have that understood by another.

We note that a multimedia scene is not reality; it is merely a reference to a referent in reality. Similarly, the output from emergence is a reference, hopefully useful to the user of the information. We use the linguistic terms "reference" and "referent" to indicate existence in the "modeled world" and the "real world," respectively. There is a danger in confusing the two (Minsky, 1988). The referenced meaning is embedded in our experience. This is similar to attribute binding using Dublin Core metadata (Hillmann, 2003), where the standard attribute name is associated with the commonly understood semantic.

The principal benefit of emergence is dealing with unanticipated situations. Units in unanticipated configurations or situations will still interact with each other in simple ways. Emergent systems, ideally, take care of themselves, without needing intervention or anticipation on the part of the system architect (Staab 2002). However, the main advantage of emergent semantics is also its greatest flaw. As well as dealing with unanticipated situations, it can also produce unanticipated results. We cannot control the outcomes. They might be useful, trivial or useless, or — in the worst case — misleading. However, we can constrain the scope of output by constraining the inputs and the ground truths. We can also ask for multiple interpretations. Sometimes, a structure is better understood if one can appreciate the other forms it can take.

In the next section, we state the requirements of emergent semantics. This will be followed by a description of existing research. In the last section, we identify gaps in the research, and suggest future directions.

EMERGENT SYSTEMS

Both complete order (regularity) and complete chaos (randomness) are very simple. Complexity occurs between the two, at a place known as "the edge of chaos" (Langton, 1990). Emergence results in complex systems, forming spontane-

ously from the interactions of many simple units. In nature, emergence is typically expressed in self-assembly, such as (micro-level) crystal formation and (macro-level) weather systems. These systems form naturally without centralized control. Similarly, emergence is useful in computer systems, when centralized control is impractical. The resources needed in these systems are primarily simple building blocks capable of interacting with each other and their environment (Holland, 2000). However, we are not interested in all possible complex systems that may form. We are interested in systems that might form useful semantic structures. We need to set up environments where the emergence is likely to result in complex semantic representation or expression (Whitesides & Grzybowski, 2003; Crutchfield, 1993; Potgeiter & Bishop, 2002). It is therefore necessary to understand the characteristics and issues involved in emergent information systems.

We lead our discussion through the example of an ant colony. An ant colony is comprised, primarily, of many small units known as ants. Each ant can only do simple tasks; for example, walk, carry, lay a pheromone trail, follow a trail, and so forth. However, the colony is sophisticated enough to thoroughly explore and manage its environment. Several characteristics of emergent systems are demonstrated in the ant colony metaphor: interaction, synthesis and self-organization. The main emergent phenomenon is self-organization, expressed in specialized ants being where the colony needs them, when appropriate. These ants and others, the ant interactions, the synthesis and self-organization, compose the ant colony. See Bonabeau and Theraulaz (2000) for more details.

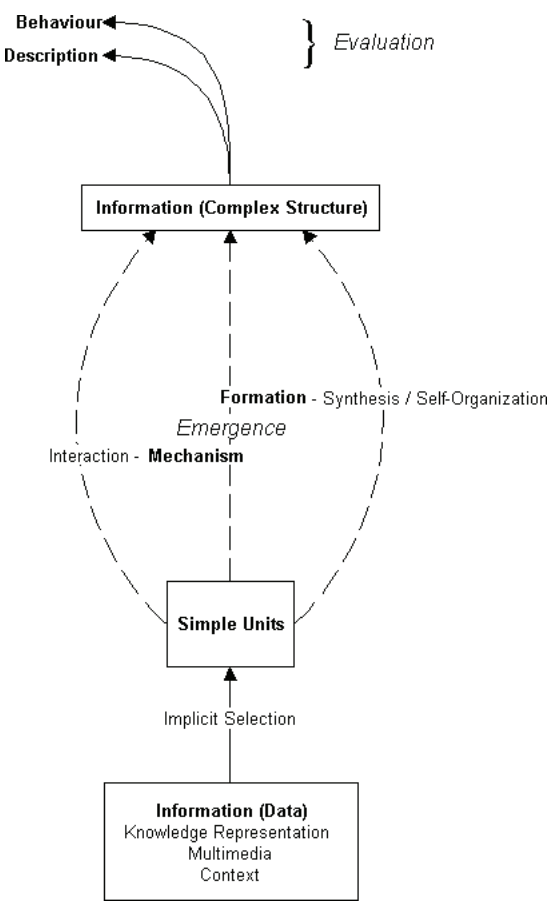
This section describes the characteristics and practical issues of emergent systems. They constitute our requirements. These include, but are not limited to, interaction, synthesis, self-organization, knowledge representation, context and evaluation.

Characteristics of Emergent Systems

The requirements of emergent semantic systems (Figure 1) can be seen from three main perspectives: information, mechanism and formation. Information is what we ultimately want (the result of the emergence). It can also be what we need (a source of units for interaction). Some of this information is implicit in the multimedia and context. Other information is either implicit or explicit in stored knowledge.

It might seem that nothing particularly useful happens at the scope of two units interacting. However, widening our field of view to take in the interaction of many units, we should see synthesis

Figure 1. Emergent semantic systems



of complex semantic units, and eventually self-organization of the unit population into semantic structures. These semantic structures can then be used to address the information need and augment the knowledge used by initial interaction.

Context influences the emergent structures. We need mechanisms which enable different interactions to occur depending on the context. These mechanisms also need to implicitly select, from the data, the salient units for each situation; either that or cause the interaction of those units, to have a greater effect.

Information

If humans are to evaluate the emergence, they must either observe system behaviour (phenotype) or a knowledge representation (genotype). The system representation must be translatable to terms a human can understand, or to an intermediate representation that can provide interaction. Typically, for this to be possible, the domain needs to be well known. Unanticipated events might not be translated well. Though we deal with the unanticipated, we must communicate in terms of the familiar. Emergence must be in terms of the system being interpreted. Otherwise we run the risk of infinite regression (Crutchfield, 1993). The environment, context and user should be included as part of the system. We need semantic structures, which contain the result of emergence, to be part of the system.

Context will either determine which of the many interpretations are appropriate or constrain the interpretation formation. Context is taken mainly from the user or from the application domain. Spatial and temporal positioning of features can also provide context, depending on the domain. The significance of specialized information, such as geographical position or time point, would be part of application domains such as fire fighting or astronomy. It is known in film theory as the Kuleshov effect. Reordering shots in a scene affects interpretation (Davis, Dorai, & Nack, 2003). Context supplies the system with constraints on relationships between entities. It can also affect the granularity and form of semantic output: classification, labelled multimedia objects, metadata, semantic networks, natural language description, or system behaviour. Different people will want to know different things.

Mechanism

The defining characteristic of useful emergent systems is that simple units can interact to provide complex and useful structures. Interaction[1] is the notion that units[2] in the system will interact to form a combined entity, which has properties that no unit has separately. The interaction is significant. Examining the units in isolation will not completely explain the properties of the whole. For two or more units to interact some mechanism must exist which enables them to interact. The mere presence of two salient units doesn't mean that they are able to interact.

Before we can reap the benefits of units interacting, we need units. These units might be implied by the data. Explicit selection of units by a central controller would not be part of an emergent process. Emergence involves implicit selection of the right units to interact. The environment should make it likely for salient units to interact. Possibly all units interact, with the salient units interacting more. In different contexts, different units will be the salient units. The context should change which units are more likely to interact, or the significance of their interaction.

Formation

"Bridge laws, linking micro and macro properties, are emergent laws if they are not semantically implied by initial micro conditions and micro laws" (McLaughlin, 2001).

Synthesis involves a group of units composing a recognisable whole. Most systems do analysis

— which involves top-down reduction and a control structure performing analysis. Synthesis is essentially the interaction mechanism seen at another level or from a different perspective. A benefit of emergence is that the system designer is freed from having to anticipate everything. Synthesis involves bottom-up emergence, which results in a complex structure. The unanticipated interaction of simple units might carry out an unanticipated and complex task. We lessen the need for a high-level control structure that tries to anticipate all possible future scenarios. Boids (Reynolds, 1987) synthesizes flocking behaviour in a population of simple units. Each unit in the flock follows simple laws, knowing only how to interact with its closest neighbours. The knowledge of forming a flock isn't stored in any unit in the flock. Unanticipated obstacles are avoided by the whole flock, which reforms if split up.

Self-Organization involves a population of units which appear to determine their own collective form and processes. "Self-assembly is the autonomous organization of components into patterns or structures without human intervention" (Whitaker, 2003; Whitesides & Grzybowski, 2003). Similar though less complex, self-organization occurs in *artificial life* (Waldrop, 1992). It attempts to mimic biological systems by capturing an abstract model of evolution. Organisms have genes, which specify simple attributes or behaviours. Populations of organisms interact to produce complex systems.

ISSUES

Evaluation

The 'Semantic Gap' is industry jargon for the gap between sensory information and the complex model in a human's mind. The same sensory information provides some of the units which participate in computational emergence. The semantic structures, which are formed, are the system's complex model. Since emergence is not something controlled, we cannot make sure that the system's complex model will be the same as the human's complex model. The ant colony is not controlled, though we consider it successful. If the ant colony self-organized in a different way, we might consider that structure successful as well. There may be many acceptable, emergent semantic structures. We need to know whether the semantic emergence is appropriate, to either the user or task. Therefore, we need to evaluate the emergence, either through direct communication of the semantic structure or though system behaviour.

Scalability

This notion of scale is slightly different to traditional notions. We can scale with respect to domain and richness of data. Most approaches for semantics constrain the domain of knowledge, such that the constraints themselves provide ground truths. If a system tries to cater for more domains, it loses some of the ground truths. Richness of data refers to numbers of units and types of units available. If the amount of data is too small, we might not have enough interaction to create a meaningful structure. A higher amount of data increases the number of units and types available. Unit pairings increase exponentially with increasing units. A system where all units try to interact with all other units might stress the process power of the system. A system without all possible interactions might miss the salient interactions. It is also uncertain whether increasing data richness will lead to finer granularity of semantic structure or lesser ability to settle on a stable structure.

Augmentation

Especially for iterative processes, it might be useful to incrementally add knowledge back to the system. The danger here is that in order to reapply what has been "learned" the system will

have to recognise situations which have occurred before with different sensory characteristics; for example, two pictures of the same situation taken from different angles.

CURRENT RESEARCH

Having described the requirements in abstract, we will now describe the tools and techniques which address the requirements.

Information

Knowledge representation, for emergence, includes ontology, metadata and genetic algorithm strings. The use of templates and grammars, which can communicate semantics in terms of the media, aren't emergent techniques as their semantic structure is predefined and the multimedia content anticipated.

Metadata (data about data) can be used as an alternative semantic description of multimedia content. MPEG-7 has a description stream, which is associated with the multimedia stream by using temporal operators. The description resides with the data. However, it is difficult in advance to provide metadata for every possible future interpretation of an event. The metadata can instead be derived from emergent semantic structures. If descriptions are needed, natural language can be derived from predicates associated with modeled concepts (Kojima, Tamura, & Fukunaga, 2002).

Classically, ontology is the study of being. The computer industry uses the term to refer to fact bases, repositories of properties and relations between objects, and semantic networks (such as Princeton's WordNet). Some ontology is used for reference, with multimedia objects or direct sensory inputs being used to index the ontology (Hoogs, 2001; Kuipers, 2000). Other ontology attempts to capture how humans communicate their own cognitive structures. The Semantic Web attempts to use ontology to access the semantics implicit in human communication (Maedche, 2002). Semantic networks consist of a skeleton of low-level data which can be augmented by adding semantic annotation nodes (Nack, 2002). The low-level data consists of the multimedia or ground truths, which can act as units in an emergent system. The annotation nodes can contain the results of emergence, and they are not permanent. This has the advantage of providing metadata-like properties, which can also be changed for different contexts.

In genetic algorithms, knowledge representation (genotype) lies in evolving strings. The strings can contain units and the operators that act on them (Gero & Ding, 1997). The genotypes evolve over several generations, with the successful[3] genes selected to generate the next generation. Knowledge and context are acquired across generations.

Context is taken from the domain, the data instances, or the user. A user's context is mainly taken from their interaction history. Their personal history and current mental state are harder to measure. The user's role during context gathering can be active (direct manipulation) or passive (observation).

Direct Manipulation

The user can actively communicate context to the system. Santini, Gupta, and Jain (2001) ask their users to organize images in a database. They use the example of a portrait. If the portrait is in a cluster of paintings, then the semantic is "painting." If it is in a cluster of people, the semantic is "people" or "faces". The same image can be a reference to different referents, which can be intangible ideas as well as tangible objects. CollageMachine (Kerne, 2002) is a Web browsing tool which tries to predict user browsing intentions. The system tries to predict possible lines of user inquiry and selects multimedia components of those to display. Reorganization of those com-

ponents, by the user, is used by the system to adjust its model.

Observation

The context of a multimedia instance is taken from past and future subjects of user attention. The entire path taken, or group formed, by a user provides an interpretation for an individual node. The whole provides the context for the part. Emergence depends on what the user thinks the data are, though the user does not need to know how they draw conclusions from observing the data. The emergence of semantics can be made by observing human and machine agent interaction (Staab, 2002). Context, at each point in the user's path, is supplied by their navigation (Grosky, Sreenath, & Fotouhi, 2002). The user's interpretation can be different from the authors' intentions. The Web is considered a directed graph (nodes: Web pages, edges: links). Adjacent nodes are considered likely to have similar semantics, though attempts are made to detect points of interest change. The meaning of a Web page (and multimedia instances in general) emerges through use and observation.

Grouping

In order to model what humans can observe, it is often helpful to model human vision. In computer vision, grouping algorithms (based on human vision) are used to form higher-level structures from units within an image (Engbers & Smeulders, 2003). An algorithm can be an emergent technique if it can adapt dynamically to context.

Context can come from sources other than users. Multiple media can be associated with data events to help in disambiguating semantics (Nakamura & Kanade, 1997). Context in genetic algorithms is sensed over many generations, if one interprets that better performance in the environment is response to context. The domain, in schemata agreement, is partly defined by the parties involved.

Mechanism

Mechanisms of automatic, implicit unit selection and interaction are yet to be developed for semantic emergence. This is a gap in the literature that will need to be filled. Current mechanisms involve the user as a unit. The semantics emerge through interaction of the user's own context with multimedia components (Santini & Jain, 1999; Kerne, 2002). The user decides which things interact, either actively or passively. In genetic algorithms, fitness functions decide how gene strings evolve (Gero & Ding, 1997). Genetic algorithms can be used to lessen the implicit selection problem by reducing the search spaces of how units interact, which units interact, and which things are considered units.

A similar situation arises with evaluation of emerged semantics. The current thinking is that humans are needed to evaluate accuracy or reasonableness. In genetic algorithms, the representation (genotype) can be evaluated indirectly by testing the phenotype (expression).

Simply having all the necessary sensory (and other) information present will not necessarily result in interaction occurring. Information from ontology could be used in decision making, or in suggesting other units for interaction. Explicitly identifying units for interaction might be a practical nonemergent step. Units can be feature patterns rather than individual features (Fan, Gao, Luo, & Hacid, 2003). Templates can be used to search for units suggested by the ontology. Well-known video structures can be used to locate salient units within video sequences (Russell, 2000; Dorai & Venkatesh, 2001; Venkatesh & Dorai, 2001). Data can provide context by affecting the perception or emotions of the observer. Emotions can be referents. Low-level units, such as tempo and colour, in the multimedia instance, act as symbols which reference them.

Formation

In genetic algorithms, synthesis occurs between generations. The genotype is self-organizing. With direct manipulation and user observation, synthesis and organization come in the form of users putting things together.

In schemata agreement, region synthesis leads to self-organization. It is designed to be adaptable to unfamiliar schemata. Agreement can be used to capture relationships later. Emergence occurs as pairs of nodes in decentralised P2P (peer-to-peer) systems attempt to form global semantic agreements by mapping their respective schemata (Aberer, Cudre-Mauroux, & Hauswirth, 2003). Regions of similar property emerge as nodes in the network are connected pair wise, and as other nodes link to them (Langley, 2001). Unfortunately, this work does not deal with multimedia. There has been recent interest in combining the areas of multimedia, data mining and knowledge discovery. However, the semantics here are not emergent. There is also data mining research into multimedia using Self-Organizing Maps (SOM), but this is not concerned with semantics (Petrushin, Kao, & Khan, 2003; Simoff & Zaiane, 2000).

CONCLUSION

A major difficulty in researching the concept of emergent semantics in multimedia is that there are no complete systems integrating the various techniques. While there is work in knowledge representation, with respect to both semantics and multimedia, to the best of our knowledge, there's very little in interaction, synthesis and self-organization. There is the work on schemata agreement (nonmultimedia) and some work on Self-Organizing Maps (nonsemantic), but nothing combining them. The little that has been done involves users to provide context and genetic algorithms to reduce problem spaces.

One of the gaps to be filled is developing interaction mechanisms, which enable possibly unanticipated data to interact with each other and their environment. Even if we can trust the process, we are still dependent on its inputs — the simple units that interact. The set of units needs to be sufficiently rich to enable acceptable emergence. Ideally, salient features (even patterns) should naturally select themselves during emergence, though this may require participation of all units, placing a high computational load on the system. Part of the problem, for emergence techniques, is that the simple interactions must occur in parallel, and in numbers great enough to realise self-organization. The future will probably have more miniaturized systems, capable of true parallelism in quantum computers. A cubic millimetre of the brain holds the equivalent of 4 km of axonal wiring (Koch, 2001). Perhaps greater parallelism will permit interaction of all available units.

There is motivation for research into nonverbal computing, where the users are illiterate (Jain, 2003). Without user ability to issue and access abstract concepts, the concepts must be inferred. Experiential computing (Jain, 2003; Sridharan, Sundaram, & Rikasis, 2003) allows users to interact with the system environment, without having to build a mental model of the environment. They seek a symbiosis formed from human and machine, taking advantage of their respective strengths. These systems are insight facilitators. They help us make sense of our own context by engaging our senses directly, as opposed to being confronted by an abstract description. Experiential computing, while in its infancy now, might in the future enable implicit relevance feedback. The user's interactions with the system could cause both emergence and verification of semantics.

REFERENCES

Aberer, K., Cudre-Mauroux, P., & Hauswirth, M. (2003). *T*he chatty Web: Emergent semantics through gossiping. Paper presented at the *WWW2003*, Budapest, Hungary.

Bonabeau, E., & Theraulaz, G. (2000). Swarm smarts. *Scientific American, 282*(3), 54-61.

Crutchfield, J. P. (1993). The calculi of emergence. Paper presented at the *Complex Systems - from Complex Dynamics to Artificial Reality*, Numazu, Japan.

Davis, M., Dorai, C., & Nack, F. (2003). *Understanding media semantics.* Berkeley, CA: ACM Multimedia 2003 Tutorial.

Dorai, C., & Venkatesh, S. (2001, September 10-12). Bridging the semantic gap in content management systems: Computational media aesthetics. Paper presented at the *COSIGN 2001: Computational Semiotics* (pp. 94-99), CWI Amsterdam.

Engbers, E. A., & Smeulders, A. W. M. (2003). Design considerations for generic grouping in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(4), 445-457.

Fan, J., Gao, Y., Luo, H., & Hacid, M.-S. (2003). A novel framework for semantic image classification and benchmark. Paper presented at the *ACM SIGKDD*, Washington, DC.

Gero, J. S., & Ding, L. (1997). Learning emergent style using and evolutionary approach. In B.Varma & X. Yao (Eds.), paper presented at the *ICCIMA* (pp. 171-175), Gold Coast, Australia.

Grosky, W. I., Sreenath, D. V., & Fotouhi, F. (2002). Emergent semantics and the multimedia semantic Web. *SIGMOD Record, 31*(4), 54-58.

Hillmann, D. (2003). *Using Dublin core.* Retrieved February 16, 2004, from *http://dublincore.org/documents/usageguide/*

Holland, J. H. (2000). *Emergence: From chaos to order* (1st ed.). Oxford: Oxford University Press.

Hoogs. (2001, 10-12 October). Multi-modal fusion for video understanding. Paper presented at the *30th Applied Imagery Pattern Recognition Workshop* (pp. 103-108), Washington, DC.

Jain, R. (2003). Folk computing. *Communications of the ACM, 46*(3), 27-29.

Kerne, A. (2002). Concept-context-design: A creative model for the development of interactivity. Paper presented at the *Creativity and Cognition*, Vol. 4 (pp. 92-122), Loughborough, UK.

Koch, C. (2001). Computing in single neurons. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 174-176). Cambridge, MA: MIT Press.

Kojima, A., Tamura, T., & Fukunaga, K. (2002). Natural language description of human activities from video images based on concept hierarchy of actions. *International Journal of Computer Vision, 150*(2), 171-184.

Kuipers, B. J. (2000). The spatial semantic hierarchy. *Artificial Intelligence, 119*, 191-233.

Langley, A. (2001). Freenet. In A. Oram (Ed.), *Peer-to-peer: Harnessing the benefits of a disruptive technology* (pp. 123-132). Sebastopol, CA: O'Reilly.

Langton, C. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D, 42*(1-3), 12-37.

Maedche. (2002). Emergent semantics for ontologies. *IEEE Intelligent Systems, 17*(1), 85-86.

McLaughlin, B. P. (2001). Emergentism. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 267-269). Cambridge, MA: MIT Press.

Minsky, M. L. (1988). *The society of mind* (1st ed.). New York: Touchstone (Simon & Schuster).

Nack, F. (2002). The future of media computing. In S. Venkatesh & C. Dorai (Eds.), *Media computing* (159-196). Boston: Kluwer.

Nakamura, Y., & Kanade, T. (1997, November). Spotting by association in news video. Paper presented at the *Fifth ACM International Multimedia Conference* (pp. 393-401) Seattle, Washington.

OED. (2003). *Oxford English Dictionary.* Retrieved February 2004, from *dictionary.oed.com/entrance.dtl*

Petrushin, V. A., Kao, A., & Khan, L. (2003). *The Fourth International Workshop on Multimedia Data Mining*, MDM/KDD 2003. Vol. 6(1). (pp. 106-108).

Potgeiter, A., & Bishop, J. (2002). *Complex adaptive systems, emergence and engineering: The basics.* Retrieved February 20, 2004, from *http://people.cs.uct.ac.za/~yng/Emergence.pdf*

Reynolds, C. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*(4), 25-34.

Russell, D. (2000). A design pattern-based video summarization technique. Paper presented at the *Proceedings of the 33rd Hawaii International Conference on System Sciences* (p. 3048).

Santini, S., Gupta, A., & Jain, R. (2001). Emergent semantics through interaction in image databases. *IEEE Transactions on Knowledge and Data Engineering, 13*(3), 337-351.

Santini, S., & Jain, R. (1999, Jan). Interfaces for emergent semantics in multimedia databases. Paper presented at the *SPIE,* San Jose, California.

Simoff, S. J., & Zaiane, O. R. (2000). Report on MDM/KDD2000: The First International Workshop on Multimedia Data Mining. *SIGKDD Explorations, 2*(2), 103-105.

Sridharan, H., Sundaram, H., & Rikasis, T. (2003, November 7). Computational models for experiences in the arts, and multimedia. Paper presented at the *ACM Multimedia 2003, First ACM Workshop on Experiential Teleprescence*, Berkeley, CA, USA.

Staab, S. (2002). Emergent semantics. *IEEE Intelligent Systems, 17*(1), 78-79.

Venkatesh, S., & Dorai, C. (2001). Computational media aesthetics: Finding meaning beautiful. *IEEE Multimedia*, 10-12.

Waldrop, M. M. (1992). Life at the edge of chaos. In *Complexity*, (pp. 198-240). New York: Touchstone (Simon & Schuster).

Whitaker, R. (2003). Self-organization, autopoiesis and enterprises. Retrieved February 3, 2005, from *http://www.acm.org/sigois/auto/Main.html*

Whitesides, G. M., & Grzybowski, B. (2003). Self-assembly at all scales. *Science, 295*, 2418-2421.

ENDNOTES

1. The action or influence of persons or things on each other (OED, 2003).
2. The user can also be considered a unit.
3. According to a fitness function, which measures the phenotype (the string's expression or behaviour).

This work was previously published in Managing Multimedia Semantics, edited by U. Srinivasan and S. Nepal, pp. 351-362, copyright 2005 by IRM Press (an imprint of IGI Global).

Section 2

Development and Design Methodologies

This section provides in-depth coverage of conceptual architectures, frameworks and methodologies related to the design and implementation of multimedia technologies. Throughout these contributions, research fundamentals in the discipline are presented and discussed. From broad examinations to specific discussions on electronic tools, the research found within this section spans the discipline while also offering detailed, specific discussions. Basic designs, as well as abstract developments, are explained within these chapters, and frameworks for designing successful multimedia interfaces, applications, and even environments are discussed.

Chapter 2.1
Content-Based Multimedia Retrieval

Chia-Hung Wei
University of Warwick, UK

Chang-Tsun Li
University of Warwick, UK

INTRODUCTION

In the past decade, there has been rapid growth in the use of digital media such as images, video, and audio. As the use of digital media increases, effective retrieval and management techniques become more important. Such techniques are required to facilitate the effective searching and browsing of large multimedia databases.

Before the emergence of content-based retrieval, media was annotated with text, allowing the media to be accessed by text-based searching (Feng et al., 2003). Through textual description, media can be managed, based on the classification of subject or semantics. This hierarchical structure allows users to easily navigate and browse, and can search using standard Boolean queries. However, with the emergence of massive multimedia databases, the traditional text-based search suffers from the following limitations (Djeraba, 2003; Shah et al., 2004):

- Manual annotations require too much time and are expensive to implement. As the number of media in a databases grows, the difficulty finding desired information increases. It becomes infeasible to manually annotate all attributes of the media content. Annotating a 60-minute video containing more than 100,000 images consumes a vast amount of time and expense.

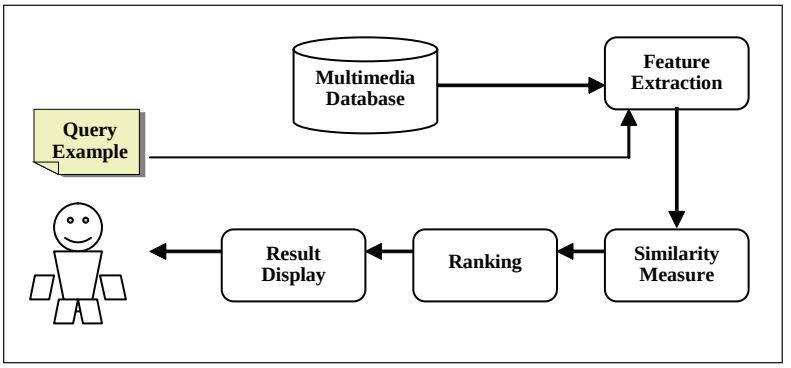
- Manual annotations fail to deal with the discrepancy of subjective perception. The phrase "a picture is worth a thousand words" implies that the textual description is not sufficient for depicting subjective perception. Capturing all concepts, thoughts, and feelings for the content of any media is almost impossible.
- Some media contents are difficult to describe concretely in words. For example, a piece of melody without lyrics or an irregular organic shape cannot be expressed easily in textual form, but people expect to search media with similar contents based on examples they provide. In an attempt to overcome these difficulties, content-based retrieval employs content information to automatically index data with minimal human intervention.

APPLICATIONS

Content-based retrieval has been proposed by different communities for various applications. These include:

- **Medical Diagnosis:** The amount of digital medical images used in hospitals has increased tremendously. As images with the similar pathology-bearing regions can be found and interpreted, those images can be applied to aid diagnosis for image-based reasoning. For example, Wei & Li (2004) proposed a general framework for content-based medical image retrieval and constructed a retrieval system for locating digital mammograms with similar pathological parts.
- **Intellectual Property:** Trademark image registration has applied content-based retrieval techniques to compare a new candidate mark with existing marks to ensure that there is no repetition. Copyright protection also can benefit from content-based retrieval, as copyright owners are able to search and identify unauthorized copies of images on the Internet. For example, Wang & Chen (2002) developed a content-based system using hit statistics to retrieve trademarks.
- **Broadcasting Archives:** Every day, broadcasting companies produce a lot of audiovisual data. To deal with these large archives, which can contain millions of hours of video and audio data, content-based retrieval techniques are used to annotate their contents and summarize the audiovisual data to drastically reduce the volume of raw footage. For example, Yang et al. (2003) developed a content-based video retrieval system to support personalized news retrieval.

Figure 1. A conceptual architecture for content-based retrieval



- **Information Searching on the Internet:** A large amount of media has been made available for retrieval on the Internet. Existing search engines mainly perform text-based retrieval. To access the various media on the Internet, content-based search engines can assist users in searching the information with the most similar contents based on queries. For example, Hong & Nah (2004) designed an XML scheme to enable content-based image retrieval on the Internet.

DESIGN OF CONTENT-BASED RETRIEVAL SYSTEMS

Before discussing design issues, a conceptual architecture for content-based retrieval is introduced and illustrated in Figure 1.

Content-based retrieval uses the contents of multimedia to represent and index the data (Wei & Li, 2004). In typical content-based retrieval systems, the contents of the media in the database are extracted and described by multi-dimensional feature vectors, also called descriptors. The feature vectors of the media constitute a feature dataset. To retrieve desired data, users submit query examples to the retrieval system. The system then represents these examples with feature vectors. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature dataset are then computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way to search the media database. Finally, the system ranks the search results and then returns the top search results that are the most similar to the query examples.

For the design of content-based retrieval systems, a designer needs to consider four aspects: feature extraction and representation, dimension reduction of feature, indexing, and query specifications, which will be introduced in the following sections.

FEATURE EXTRACTION AND REPRESENTATION

Representation of media needs to consider which features are most useful for representing the contents of media and which approaches can effectively code the attributes of the media. The features are typically extracted off-line so that efficient computation is not a significant issue, but large collections still need a longer time to compute the features. Features of media content can be classified into low-level and high-level features.

Low-Level Features

Low-level features such as object motion, color, shape, texture, loudness, power spectrum, bandwidth, and pitch are extracted directly from media in the database (Djeraba, 2002). Features at this level are objectively derived from the media rather than referring to any external semantics. Features extracted at this level can answer queries such as "finding images with more than 20% distribution in blue and green color," which might retrieve several images with blue sky and green grass (see Picture 1). Many effective approaches to low-level feature extraction have been developed for various purposes (Feng et al., 2003; Guan et al., 2001).

High-Level Features

High-level features are also called semantic features. Features such as timbre, rhythm, instruments, and events involve different degrees of semantics contained in the media. High-level features are supposed to deal with semantic queries (e.g.,"finding a picture of water" or "searching for Mona Lisa Smile"). The latter query contains higher-degree semantics than the former. As water in images displays the homogeneous texture represented in low-level features, such a query is easier to process. To retrieve the latter query, the

retrieval system requires prior knowledge that can identify that Mona Lisa is a woman, who is a specific character rather than any other woman in a painting.

The difficulty in processing high-level queries arises from external knowledge with the description of low-level features, known as the semantic gap. The retrieval process requires a translation mechanism that can convert the query of "Mona Lisa Smile" into low-level features. Two possible solutions have been proposed to minimize the semantic gap (Marques & Furht, 2002). The first is automatic metadata generation to the media. Automatic annotation still involves the semantic concept and requires different schemes for various media (Jeon et al., 2003). The second uses relevance feedback to allow the retrieval system to learn and understand the semantic context of a query operation. Relevance feedback will be discussed in the Relevance Feedback section.

Picture 1. There are more than 20% distributions in blue and green color in this picture

DIMENSION REDUCTION OF FEATURE VECTOR

Many multimedia databases contain large numbers of features that are used to analyze and query the database. Such a feature-vector set is considered as high dimensionality. For example, Tieu & Viola (2004) used over 10,000 features of images, each describing a local pattern. High dimensionality causes the "curse of dimension" problem, where the complexity and computational cost of the query increases exponentially with the number of dimensions (Egecioglu et al., 2004). Dimension reduction is a popular technique to overcome this problem and support efficient retrieval in large-scale databases. However, there is a tradeoff between the efficiency obtained through dimension reduction and the completeness obtained through the information extracted. If each data is represented by a smaller number of dimensions, the speed of retrieval is increased. However, some information may be lost. One of the most widely used techniques in multimedia retrieval is Principal Component Analysis (PCA). PCA is used to transform the original data of high dimensionality into a new coordinate system with low dimensionality by finding data with high discriminating power. The new coordinate system removes the redundant data and the new set of data may better represent the essential information. Shyu et al. (2003) presented an image database retrieval framework and applied PCA to reduce the image feature vectors.

INDEXING

The retrieval system typically contains two mechanisms: similarity measurement and multidimensional indexing. Similarity measurement is used to find the most similar objects. Multidimensional indexing is used to accelerate the query performance in the search process.

Similarity Measurement

To measure the similarity, the general approach is to represent the data features as multi-dimensional points and then to calculate the distances between the corresponding multi-dimensional points (Feng et al., 2003). Selection of metrics has a direct impact on the performance of a retrieval system. Euclidean distance is the most common metric used to measure the distance between two points in multi-dimensional space (Qian et al., 2004). However, for some applications, Euclidean distance is not compatible with the human perceived similarity. A number of metrics (e.g., Mahalanobis Distance, Minkowski-Form Distance, Earth Mover's Distance, and Proportional Transportation Distance) have been proposed for specific purposes. Typke et al. (2003) investigated several similarity metrics and found that Proportional Transportation Distance fairly reflected melodic similarity.

Multi-Dimensional Indexing

Retrieval of the media is usually based not only on the value of certain attributes, but also on the location of a feature vector in the feature space (Fonseca & Jorge, 2003). In addition, a retrieval query on a database of multimedia with multi-dimensional feature vectors usually requires fast execution of search operations. To support such search operations, an appropriate multi-dimensional access method has to be used for indexing the reduced but still high dimensional feature vectors. Popular multi-dimensional indexing methods include R-tree (Guttman, 1984) and R*-tree (Beckmann et al., 1990). These multi-dimensional indexing methods perform well with a limit of up to 20 dimensions. Lo & Chen (2002) proposed an approach to transform music into numeric forms and developed an index structure based on R-tree for effective retrieval.

QUERY SPECIFICATIONS

Querying is used to search for a set of results with similar content to the specified examples. Based on the type of media, queries in content-based retrieval systems can be designed for several modes (e.g., query by sketch, query by painting [for video and image], query by singing [for audio], and query by example). In the querying process, users may be required to interact with the system in order to provide relevance feedback, a technique that allows users to grade the search results in terms of their relevance. This section will describe the typical query by example mode and discuss the relevance feedback.

Query by Example

Queries in multimedia retrieval systems are traditionally performed by using an example or series of examples. The task of the system is to determine which candidates are the most similar to the given example. This design is generally termed Query By Example (QBE) mode. The interaction starts with an initial selection of candidates. The initial selection can be randomly selected candidates or meaningful representatives selected according to specific rules. Subsequently, the user can select one of the candidates as an example, and the system will return those results that are most similar to the example. However, the success of the query in this approach heavily depends on the initial set of candidates. A problem exists in how to formulate the initial panel of candidates that contains at least one relevant candidate. This limitation has been defined as page zero problem (La Cascia et al., 1998). To overcome this problem, various solutions have been proposed for specific applications. For example, Sivic and Zisserman (2004) proposed a method that measures the reoccurrence of spatial configurations of viewpoint invariant features to obtain the principal objects, characters, and scenes, which can be used as entry points for visual search.

Relevance Feedback

Relevance feedback was originally developed for improving the effectiveness of information retrieval systems. The main idea of relevance feedback is for the system to understand the user's information needs. For a given query, the retrieval system returns initial results based on predefined similarity metrics. Then, the user is required to identify the positive examples by labeling those that are relevant to the query. The system subsequently analyzes the user's feedback using a learning algorithm and returns refined results. Two of the learning algorithms frequently used to iteratively update the weight estimation were developed by Rocchio (1971) and Rui and Huang (2002).

Although relevance feedback can contribute retrieval information to the system, two challenges still exist: (1) the number of labeled elements obtained through relevance feedback is small when compared to the number of unlabeled in the database; (2) relevance feedback iteratively updates the weight of high-level semantics but does not automatically modify the weight for the low-level features. To solve these problems, Tian et al. (2000) proposed an approach for combining unlabeled data in supervised learning to achieve better classification.

FUTURE RESEARCH ISSUES AND TRENDS

Since the 1990s, remarkable progress has been made in theoretical research and system development. However, there are still many challenging research problems. This section identifies and addresses some issues in the future research agenda.

Automatic Metadata Generation

Metadata (data about data) is the data associated with an information object for the purposes of description, administration, technical functionality, and so on. Metadata standards have been proposed to support the annotation of multimedia content. Automatic generation of annotations for multimedia involves high-level semantic representation and machine learning to ensure accuracy of annotation. Content-based retrieval techniques can be employed to generate the metadata, which can be used further by the text-based retrieval.

Establishment of Standard Evaluation Paradigm and Test-Bed

The National Institute of Standards and Technology (NIST) has developed TREC (Text REtrieval Conference) as the standard test-bed and evaluation paradigm for the information retrieval community. In response to the research needs from the video retrieval community, the TREC released a video track in 2003, which became an independent evaluation (called TRECVID) (Smeaton, 2003). In music information retrieval, a formal resolution expressing a similar need was passed in 2001, requesting a TREC-like standard test-bed and evaluation paradigm (Downie, 2003). The image retrieval community still awaits the construction and implementation of a scientifically valid evaluation framework and standard test bed.

Embedding Relevance Feedback

Multimedia contains large quantities of rich information and involves the subjectivity of human perception. The design of content-based retrieval systems has turned out to emphasize an interactive approach instead of a computer-centric approach. A user interaction approach requires human and computer to interact in refining the high-level queries. Relevance feedback is a powerful technique used for facilitating interaction between the user and the system. The research issue includes the design of the interface with regard to usability and learning algorithms, which can dynamically update the weights embedded in the query object

to model the high-level concepts and perceptual subjectivity.

Bridging the Semantic Gap

One of the main challenges in multimedia retrieval is bridging the gap between low-level representations and high-level semantics (Lew & Eakins, 2002). The semantic gap exists because low-level features are more easily computed in the system design process, but high-level queries are used at the starting point of the retrieval process. The semantic gap is not only the conversion between low-level features and high-level semantics, but it is also the understanding of contextual meaning of the query involving human knowledge and emotion. Current research intends to develop mechanisms or models that directly associate the high-level semantic objects and representation of low-level features.

CONCLUSION

The main contributions in this article were to provide a conceptual architecture for content-based multimedia retrieval, to discuss the system design issues, and to point out some potential problems in individual components. Finally, some research issues and future trends were identified and addressed.

The ideal content-based retrieval system from a user's perspective involves the semantic level. Current content-based retrieval systems generally make use of low-level features. The semantic gap has been a major obstacle for content-based retrieval. Relevance feedback is a promising technique to bridge this gap. Due to the efforts of the research community, a few systems have started to employ high-level features and are able to deal with some semantic queries. Therefore, more intelligent content-based retrieval systems can be expected in the near future.

REFERENCES

Beckmann, N., Kriegel, H.-P., Schneider, R., & Seeger, B. (1990). The R*-tree: An efficient and robust access method for points and rectangles. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Atlantic City, NJ, USA.

Djeraba, C. (2002). Content-based multimedia indexing and retrieval. *IEEE MultiMedia, 9*(2) 18-22.

Djeraba, C. (2003). Association and content-based retrieval. *IEEE Transactions on Knowledge and Data Engineering, 15*(1), 118-135.

Downie, J.S. (2003). Toward the scientific evaluation of music information retrieval systems. *Proceedings of the Fourth International Symposium on Music Information Retrieval*, Washington, D.C., USA.

Egecioglu, O., Ferhatosmanoglu, H., & Ogras, U. (2004). Dimensionality reduction and similarity computation by inner-product approximations. *IEEE Transactions on Knowledge and Data Engineering,16*(6), 714-726.

Feng, D., Siu, W.C., & Zhang, H.J. (Eds.). (2003). *Multimedia information retrieval and management:Technological fundamentals and applications.* Berlin: Springer.

Fonseca, M.J., & Jorge, J.A. (2003). Indexing highdimensional data for content-based retrieval in large database. *Proceedings of the Eighth International Conference on Database Systems for Advanced Applications*, Kyoto, Japan.

Guan, L., Kung S.-Y., & Larsen, J. (Eds.). (2001). *Multimedia image and video processing.* New York: CRC Press.

Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Boston, MA, USA.

Hong, S., & Nah, Y. (2004). An intelligent image retrieval system using XML. *Proceedings of the 10th International Multimedia Modelling Conference*, Brisbane, Australia.

Jeon, J., Lavrenko, V., & Manmatha, R. (2003). Automatic image annotation and retrieval using crossmedia relevance models. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada.

La Cascia, M., Sethi, S., & Sclaroff, S. (1998). Combining textual and visual cues for content-based image retrieval on the World Wide Web. *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, Santa Barbara, CA, USA.

Lew, M.S., Sebe, N., & Eakins, J.P. (2002). Challenges of image and video retrieval. *Proceedings of the International Conference on Image and Video Retrieval, Lecture Notes in Computer Science*, London, UK.

Lo, Y.-L., & Chen, S.-J. (2002). The numeric indexing for music data. *Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops.* Vienna, Austria.

Marques, O., & Furht, B. (2002). *Content-based image and video retrieval.* London: Kluwer.

Qian, G., Sural, S., Gu, Y., & Pramanik, S. (2004). Similarity between Euclidean and cosine angle distance for nearest neighbor queries. *Proceedings of 2004 ACM Symposium on Applied Computing*, Nicosia, Cyprus.

Rocchio, J.J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.), *The SMART retrieval system—Experiments in automatic document processing.* Englewood Cliffs, NJ: Prentice Hall.

Rui, Y., & Huang, T. (2002). Learning based relevance feedback in image retrieval. In A.C. Bovik, C.W. Chen, & D. Goldfof (Eds.), *Advances in image processing and understanding: A festschrift for Thomas S. Huang* (pp. 163-182). New York: World Scientific Publishing.

Shah, B., Raghavan, V., & Dhatric, P. (2004). Efficient and effective content-based image retrieval using space transformation. *Proceedings of the 10th International Multimedia Modelling Conference*, Brisbane, Australia.

Shyu, C.R., et al. (1999). ASSERT: A physician-in-the-loop content based retrieval system for HRCT image databases. *Computer Vision and Image Understanding, 75*(1-2), 111-132.

Sivic, J., & Zisserman, A. (2004). Video data mining using configurations of viewpoint invariant regions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA.

Smeaton, A.F., Over, P. (2003). TRECVID: Benchmarking the effectiveness of information retrieval tasks on digital video. *Proceedings of the International Conference on Image and Video Retrieval*, Urbana, IL, USA.

Tian, Q., Wu, Y., & Huang, T.S. (2000). Incorporate discriminant analysis with EM algorithm in image retrieval. *Proceedings of the IEEE International Conference on Multimedia and Expo*, New York, USA.

Tieu, K., & Viola, P. (2004). Boosting image retrieval. *International Journal of Computer Vision, 56*(1-2), 17-36.

Typke, R., Giannopoulos, P., Veltkamp, R.C. Wiering, F., & Oostrum, R.V. (2003). Using transportation distances for measuring melodic similarity. *Proceedings of the Fourth International Symposium on Music Information Retrieval*, Washington, DC, USA.

Wang, C.-C., & Chen, L.-H. (2002). Content-based color trademark retrieval system using hit statistic. *International Journal of Pattern and Artificial Intelligence,16*(5), 603-619.

Wei, C.-H., & Li, C.-T. (2004). A general framework for content-based medical image retrieval with its application to mammogram retrieval. *Proceedings of IS&T/SPIE International Symposium on Medical Imaging*, San Diego, CA, USA.

Yang, H., Chaisorn, L., Zhao, Y., Neo, S.-Y., & Chua, T.-S. (2003). VideoQA: Question answering on news video. *Proceedings of the Eleventh ACM International Conference on Multimedia*, Berkeley, CA, USA.

KEY TERMS

Boolean Query: A query that uses Boolean operators (AND, OR, and NOT) to formulate a complex condition. A Boolean query example can be "university" OR "college."

Content-Based Retrieval: An application that directly makes use of the contents of media rather than annotation inputted by the human to locate desired data in large databases.

Feature Extraction: A subject of multimedia processing that involves applying algorithms to calculate and extract some attributes for describing the media.

Query by Example: A method of searching a database using example media as search criteria. This mode allows the users to select predefined examples requiring the users to learn the use of query languages.

Relevance Feedback: A technique that requires users to identify positive results by labeling those that are relevant to the query and subsequently analyzes the user's feedback using a learning algorithm.

Semantic Gap: The difference between the high-level user perception of the data and the lower-level representation of the data used by computers. As high-level user perception involves semantics that cannot be translated directly into logic context, bridging the semantic gap is considered a challenging research problem.

Similarity Measure: A measure that compares the similarity of any two objects represented in the multi-dimensional space. The general approach is to represent the data features as multi-dimensional points and then to calculate the distances between the corresponding multi-dimensional points.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 116-122, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.2
MM4U:
A Framework for Creating Personalized Multimedia Content

Ansgar Scherp
OFFIS Research Institute, Germany

Susanne Boll
University of Oldenburg, Germany

ABSTRACT

In the Internet age and with the advent of digital multimedia information, we succumb to the possibilities that the enchanting multimedia information seems to offer, but end up almost drowning in the multimedia information: Too much information at the same time, so much information that is not suitable for the current situation of the user, too much time needed to find information that is really helpful. The multimedia material is there, but the issues of how the multimedia content is found, selected, assembled, and delivered such that it is most suitable for the user's interest and background, the user's preferred device, network connection, location, and many other settings, is far from being solved. In this chapter, we are focusing on the aspect of how to assemble and deliver personalized multimedia content to the users. We present the requirements and solutions of multimedia content modeling and multimedia content authoring as we find it today. Looking at the specific demands of creating personalized multimedia content, we come to the conclusion that a dynamic authoring process is needed in which just in time the individual multimedia content is created for a specific user or user group. We designed and implemented an extensible software framework, MM4U (short for "MultiMedia for you"), which provides generic functionality for typical tasks of a dynamic multimedia content personalization process. With such a framework at hand, an application developer can concentrate on creating personalized content in the specific

domain and at the same time is relieved from the basic task of selecting, assembling, and delivering personalized multimedia content. We present the design of the MM4U framework in detail with an emphasis for the personalized multimedia composition and illustrate the framework's usage in the context of our prototypical applications.

INTRODUCTION

Multimedia content today can be considered as the composition of different media elements, such as images and text, audio, and video, into an interactive multimedia presentation like a guided tour through our hometown Oldenburg. Features of such a presentation are typically the temporal arrangement of the media elements in the course of the presentation, the layout of the presentation, and its interaction features. *Personalization* of multimedia content means that the multimedia content is targeted at a specific person and reflects this person's individual context, specific background, interest, and knowledge, as well as the heterogeneous infrastructure of end devices to which the content is delivered and on which it is presented. The creation of personalized multimedia content means that for each intended context a custom presentation needs to be created. Hence, multimedia content personalization is the shift from one-size-fits-all to a very individual and personal one-to-one provision of multimedia content to the users. This means in the end that the multimedia content needs to be prepared for each individual user. However, if there are many different users that find themselves in very different contexts, it soon becomes obvious that a manual creation of different content for all the different user contexts is not feasible, let alone economical (see André & Rist, 1996). Instead, a *dynamic*, automated process of selecting and assembling personalized multimedia content depending on the user context seems to be reasonable.

The creation of multimedia content is typically subsumed under the notion of multimedia authoring. However, such authoring today is seen as the static creation of multimedia content. Authoring tools with graphical user interfaces (GUI) allow us to manually create content that is targeted at a specific user group. If the content created is at all "personalizable," then only within a very limited scope. First research approaches in the field of dynamic creation of personalized multimedia content are promising; however, they are often limited to certain aspects of the content personalization to the individual user. Especially when the content personalization task is more complex, these systems need to employ additional programming. As we observe that programming is needed in many cases anyway, we continue this observation consequently and propose MM4U (short for "MultiMedia for you"), a component-based object-oriented software framework to support the software development process of multimedia content personalization applications. MM4U relieves application developers from general tasks in the context of multimedia content personalization and lets them concentrate on the application domain-specific tasks. The framework's components provide generic functionality for typical tasks of the multimedia content personalization process. The design of the framework is based on a comprehensive analysis of the related approaches in the field of user profile modeling, media data modeling, multimedia composition, and multimedia presentation formats. We identify the different tasks that arise in the context of creating personalized multimedia content. The different components of the framework support these different tasks for creating user-centric multimedia content: They integrate the generic access to user profiles, media data, and associated meta data, provide support for personalized multimedia composition and layout, as well as create the context-aware multimedia presentations. With such a framework, the devel-

opment of multimedia applications becomes easier and much more efficient for different users with their different (semantic) contexts. On the basis of the MM4U framework, we are currently developing two sample applications: a personalized multimedia sightseeing tour and a personalized multimedia sports news ticker. The experiences we gain from the development of these applications give us important feedback on the evaluation and continuous redesign of the framework.

The remainder of this chapter is organized as follows: To review the notion of multimedia content authoring, in Multimedia Content Authoring Today we present the requirements of multimedia content modeling and the authoring support we find today. Setting off from this, Dynamic Authoring of Personalized Content introduces the reader to the tasks of creating personalized multimedia content and why such content can be created only in a dynamic fashion. In Related Approaches, we address the related approaches we find in the field before we present the design of our MM4U framework in The Multimedia Personalization Framework section. As the personalized creation of multimedia content is a central aspect of the framework, Creating Personalized Multimedia Content presents in detail the multimedia personalization features of the framework. Impact of Personalization to The Development of Multimedia Applications shows how the framework supports application developers and multimedia authors in their effort to create personalized multimedia content. The implementation and first prototypes are presented in Implementation and Prototypical Applications before we come to our summary and conclusion in the final section.

MULTIMEDIA CONTENT AUTHORING TODAY

In this section, we introduce the reader to current notions and techniques of multimedia content modeling and multimedia content authoring. An understanding of requirements and approaches in modeling and authoring of multimedia content is a helpful prerequisite to our goal, the dynamic creation of multimedia content. For the modeling of multimedia content we present our notion of multimedia content, documents, and presentation and describe the central characteristics of typical multimedia document models in the first subsection. For the creation of multimedia content, we give a short overview of directions in multimedia content authoring today in the second subsection.

Multimedia Content

Multimedia content today is seen as the result of a composition of different media elements (media content) in a continuous and interactive multimedia presentation. Multimedia content builds on the modeling and representation of the different media elements that form the building bricks of the composition. A multimedia document represents the composition of continuous and discrete media elements into a logically coherent multimedia unit. A multimedia document that is composed in advance to its rendering is called preorchestrated in contrast to compositions that take place just before rendering that are called live or *on-the-fly*. A multimedia document is an instantiation of a *multimedia document model* that provides the primitives to capture all aspects of a multimedia document. The power of the multimedia document model determines the degree of the multimedia functionality that documents following the model can provide. Representatives of (abstract) multimedia document models in research can be found with CMIF (Bulterman et al., 1991), Madeus (Jourdan et al., 1998), Amsterdam Hypermedia Model (Hardman, 1998; Hardman et al., 1994a), and ZYX (Boll & Klas, 2001). A *multimedia document format* or *multimedia presentation format* determines the representation of a multimedia document for the document's

exchange and rendering. Since every multimedia presentation format implicitly or explicitly follows a multimedia document model, it can also be seen as a proper means to "serialize" the multimedia document's representation for the purpose of exchange. Multimedia presentation formats can either be standardized, such as the W3C standard SMIL (Ayars et al., 2001), or proprietary such as the widespread Shockwave file format (SWF) of Macromedia (Macromedia, 2004). A *multimedia presentation* is the rendering of a multimedia document. It comprises the continuous rendering of the document in the target environment, the (pre)loading of media data, realizing the temporal course, the temporal synchronization between continuous media streams, the adaptation to different or changing presentation conditions and the interaction with the user.

Looking at the different models and formats we find, and also the terminology in the related work, there is not necessarily a clear distinction between multimedia document models and multimedia presentation formats, and also between multimedia documents and multimedia presentations. In this chapter, we distinguish the notion of multimedia document models as the definition of the abstract composition capabilities of the model; a multimedia document is an instance of this model. The term *multimedia content* or *content representation* is used to abstract from existing formats and models, and generally addresses the composition of different media elements into a coherent multimedia presentation. Independent of the actual document model or format chosen for the content, one can say that a multimedia content representation has to realize at least three central aspects: the temporal, spatial, and interactive characteristics of a multimedia presentation (Boll et al., 2000). However, as many of today's concrete multimedia presentation formats can be seen as representing both a document model and an exchange format for the final rendering of the document, we use these as an illustration of the central aspects of multimedia documents. We present an overview of these characteristics in the following listing; for a more detailed discussion on the characteristics of multimedia document models we refer the reader to (Boll et al., 2000; Boll & Klas, 2001).

- A *temporal model* describes the temporal dependencies between media elements of a multimedia document. With the temporal model, the temporal course such as the parallel presentation of two videos or the end of a video presentation on a mouse-click event can be described. One can find four types of temporal models: *point-based* temporal models, *interval-based* temporal models (Little & Ghafoor, 1993; Allen, 1983), *enhanced interval-based* temporal models that can handle time intervals of unknown duration (Duda & Keramane, 1995; Hirzalla et al., 1995; Wahl & Rothermel, 1994), *event-based* temporal models, and *script-based* realization of temporal relations. The multimedia presentation formats we find today realize different temporal models, for example, SMIL 1.0 (Bugaj et al., 1998) provides an interval-based temporal model only, while SMIL 2.0 (Ayars et al., 2001) also supports an event-based model.
- For a multimedia document not only the temporal synchronization of these elements is of interest but also their spatial positioning on the presentation media, for example, a window, and possibly the spatial relationship to other visual media elements. The positioning of a visual media element in the multimedia presentation can be expressed by the use of a *spatial model.* With it one can, for example, place one image about a caption or define the overlapping of two visual media. Besides the arrangement of media elements in the presentation, also the visual layout or design is defined in the presentation. This can range from a simple setting for background colors and fonts up

to complex visual designs and effects. In general, three approaches to spatial models can be distinguished: *absolute positioning*, *directional relations* (Papadias et al., 1995; Papadias & Sellis, 1994), and *topological relations* (Egenhofer & Franzosa, 1991). With absolute positioning we subsume both the placement of a media element at an absolute position with respect to the origin of the coordinate system and the placement at an absolute position relative to another media element. The absolute positioning of media elements can be found, for example, with Flash (Macromedia, 2004) and the Basic Language Profile of SMIL 2.0, whereas the relative positioning is realized, for example, by SMIL 2.0 and SVG 1.2 (Andersson et al., 2004b).

- A very distinct feature of a multimedia document model is the ability to specify *user interaction* in order to let a user choose between different presentation paths. Multimedia documents without user interaction are not very interesting as the course of their presentation is exactly known in advance and, hence, could be recorded as a movie. With interaction models a user can, for example, select or repeat parts of presentations, speed up a movie presentation, or change the visual appearance. For the modeling of user interaction, one can identify at least three basic types of interaction: *navigational interactions*, *design interactions*, and *movie interactions*. Navigational interaction allows the selection of one out of many presentation paths and is supported by all the considered multimedia document models and presentation formats.

Looking at existing multimedia document models and presentation formats both in industry and research, one can see that these aspects of multimedia content are implemented in two general ways: The standardized formats and research models typically implement these aspects in different variants in a structured (XML) fashion as can be found with SMIL 2.0, HTML+TIME (Schmitz et al., 1998), SVG 1.2, Madeus, and ZYX. Proprietary approaches, however, represent or program these aspects in an adequate internal model such as Macromedia's Shockwave format. Independent of the actual multimedia document model, support for the *creation* of these documents is needed — multimedia content authoring. We will look at the approaches we find in the field of multimedia content authoring in the next section.

Multimedia Authoring

While multimedia content represents the composition of different media elements into a coherent multimedia presentation, *multimedia content authoring* is the process in which this presentation is actually created. This process involves parties from different fields including media designers, computer scientists, and domain experts: Experts from the domain provide their knowledge in the field; this knowledge forms the input for the creation of a storyboard for the intended presentation. Such a storyboard forms often the basis on which creators and directors plan the implementation of the story with the respective media and with which writers, photographers, and camerapersons acquire the digital media content. Media designers edit and process the content for the targeted presentation. Finally, multimedia authors compose the preprocessed and prepared material into the final multimedia presentation. Even though we described this as a sequence of steps, the authoring process typically includes cycles. In addition, the expertise for some of the different tasks in the process can also be held by one single person. In this chapter, we are focusing on the part of the multimedia content creation process in which the prepared material is actually assembled into the final multimedia presentation.

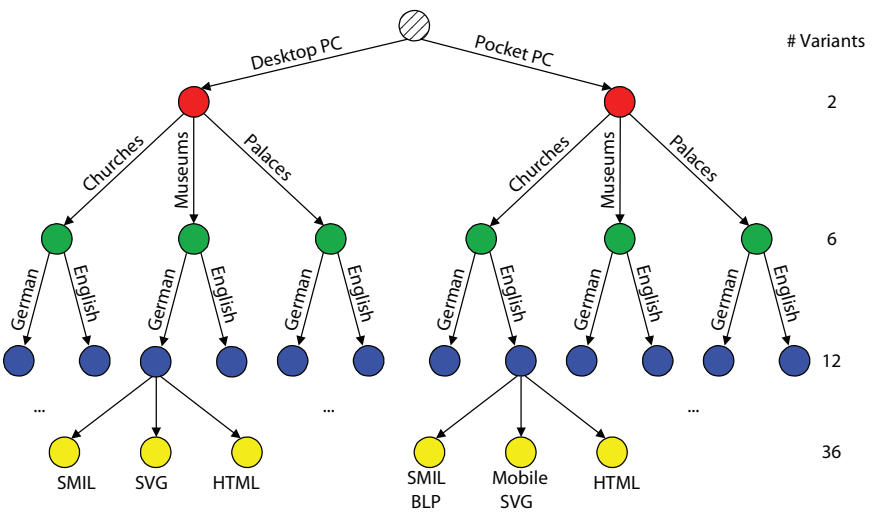
This part is typically supported by professional multimedia development programs, so-called

authoring tools or authoring software. Such tools allow the composition of media elements into an interactive multimedia presentation via a graphical user interface. The authoring tools we find here range from domain expert tools to general purpose authoring tools.

- Domain expert tools hide as much as possible the technical details of content authoring from the authors and let them concentrate on the actual creation of the multimedia content. The tools we find here are typically very specialized and targeted at a very specific domain. An example for such a tool has been developed in the context of our previous research project Cardio-OP (Klas et al., 1999) in the domain of cardiac surgery. The content created in this project is an interactive multimedia book about topics in the specialized domain of cardiac surgery. Within the project context, an easy-to-use authoring wizard was developed to allow medical doctors to easily create "pages" of a multimedia book in cardiac surgery. The Cardio-OP-Wizard guides the domain experts through the authoring process by a digital storyboard for a multimedia book on cardiac surgery. The wizard hides as much technical detail as possible.
- On the other end of the spectrum of authoring tools we find highly generalized tools such as Macromedia Director (Macromedia, 2004). These tools are independent of the domain of the intended presentation and let the authors create very sophisticated multimedia presentations. However, the authors typically need to have high expertise in using the tool. Very often programming in an integrated programming language is needed to achieve special effects or interaction patterns. Consequently, the multimedia authors need programming skills and along with this some experience in software development and software engineering.

Whereas a multimedia document model has to represent the different aspects of time, space, and interaction, multimedia authoring tools must allow the authors to actually assemble the multimedia content. However, the authors are normally experts from a specific domain. Consequently, the only authoring tools that are practicable to create

Figure 1. Example of the variation possibilities within a personalized city guide application



multimedia content for a specific domain are those that are highly specialized and easy to use.

DYNAMIC AUTHORING OF PERSONALIZED CONTENT

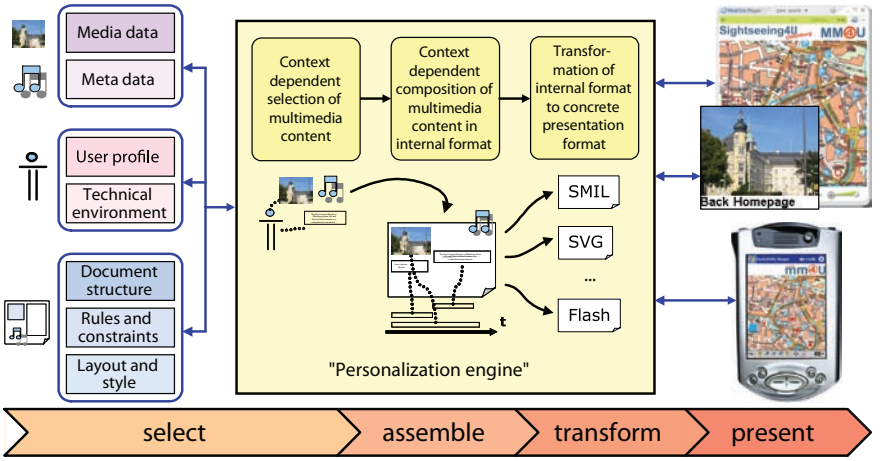
The authoring process described above so far represents a manual authoring of multimedia content, often with high effort and cost involved. Typically, the result is a multimedia presentation targeted at a certain user group in a special technical context. However, the one-size-fits-all fashion of the multimedia content created does not necessarily satisfy different users' needs. Different users may have different preferences concerning the content and also may access the content in networks on different end devices. For a wider applicability, the authored multimedia content needs to "carry" some alternatives that can be exploited to adapt the presentation to the specific preferences of the users and their technical settings. Figure 1 shows an illustration of the variation possibilities that a simple personalized city guide application can possess. The root of the tree represents the multimedia presentation for the personalized city tour. If this presentation was intended for both Desktop PC and PDA, this results in two variants of the presentation. If then some tourists are interested only in churches, museums, or palaces and would like to receive the content in either English or German, this already sums up to 12 variants. If then the multimedia content should be available in different presentation formats, the number of variation possibilities within a personalized city tour increases again. Even though different variants are not necessarily entirely different and may have overlapping content, the example is intended to illustrate that the flexibility of multimedia content to personalize to different user contexts quickly leads to an explosion of different options. And still the content can only be personalized within the flexibility range that has been anchored in the content.

From our point of view, an efficient and competitive creation of personalized multimedia content can only come from a system approach that supports the *dynamic* authoring of personalized multimedia content. A dynamic creation of such content allows for a selection and composition of just those media elements that are targeted at the user's specific interest and preferences. Generally, the dynamic authoring comprises the steps and tasks that occur also with static authoring, but with the difference that the creation process is postponed to the time when the targeted user context and the presentation is created for this specific context. To be able to efficiently create presentations for (m)any given contexts, a manual authoring of a presentation meeting the user needs is not an option; instead, a dynamic content creation is needed.

As we look into the process of *dynamic authoring* of personalized multimedia content, it is apparent that this process involves different phases and tasks. We identify the central tasks in this process that need to be supported by a suitable solution for personalized content creation.

Figure 2 depicts the general process of creating personalized multimedia content. The core of this process is an application we call *personalization engine*. The input parameters to this engine can be characterized by three groups: The first group of input parameters is the media elements with the associated meta data that constitute the content from which the personalized multimedia presentations are selected and assembled. The second group enfolds the user's personal and technical context. The user profile includes information about, for example, the user's current task, the location, and environment, like weather and loudness, his or her knowledge, goals, preferences and interests, abilities and disabilities, as well as demographic data. The technical context is described by the type of the user's end device, the hardware and software characteristics, as for example the available amount of memory and media player, as well as possible network connections and input devices.

Figure 2. General process of personalizing multimedia content



The third group of input parameters influences the general structure of the resulting personalized multimedia presentation and subsumes other preferences a user could have for the multimedia presentation.

Within the personalization engine, these input parameters are now used to author the personalized multimedia presentation. First, the personalization engine exploits all available information about the user's context and his or her end device to select by means of media meta data those media elements that are of most relevance according to the user's interests and preferences and meet the characteristics of the end device at the best. In the next step, the selected media elements are assembled and arranged by the personalization engine — again in regard to the user profile information and the characteristics of the end device — to the personalized multimedia content, represented in an *internal document model* (Scherp & Boll, 2004b). This internal document model abstracts from the different characteristics of today's multimedia presentation formats and, hence, forms the greatest common denominator of these formats. Even though our abstract model does not reflect the fancy features of some of today's multimedia presentation formats, it supports the very central multimedia features of modeling time, space, and interaction. It is designed to be efficiently transformed to the concrete syntax of the different presentation formats. For the assembly, the personalization engine uses the parameters for document structure, the layout and style parameters, and other rules and constraints that describe the structure of the personalized multimedia presentation, to determine among others the temporal course and spatial layout of the presentation. The center of Figure 2 sketches this temporal and spatial arrangement of selected media elements over time in a spatial layout following the document structure and other preferences. Only then in the transformation phase, the multimedia content in the internal document model is transformed to a concrete presentation format. Finally, the just generated personalized multimedia presentation is rendered and displayed by the actual end device.

RELATED APPROACHES

In this section we present the related approaches in the field of personalized multimedia content creation. We first discuss the creation of personalizable multimedia content with today's authoring environments before we come to research approaches that address a *dynamic* composition of adapted or personalized multimedia content.

Multimedia authoring tools like Macromedia Director (Macromedia, 2004) today require high expertise from their users and create multimedia presentations that are targeted only at a specific user or user group. Everything "personalizable" needs to be programmed or scripted within the tool's programming language. Early work in the field of creating advanced hypermedia and multimedia documents can be found, for example, with the Amsterdam Hypermedia Model (Hardman, 1998; Hardman et al., 1994b) and the authoring system CMIFed (van Rossum, 1993; Hardman et al., 1994a) as well as with the ZYX (Boll & Klas, 2001) multimedia document model and a domain-specific authoring wizard (Klas et al., 1999). In the field of standardized models, the declarative description of multimedia documents with SMIL allows for the specification of adaptive multimedia presentations by defining presentation alternatives by using the "switch" element. A manual authoring of such documents that are adaptable to many different contexts is too complex; also the existing authoring tools such as GR*i*NS editor for SMIL from Oratrix (Oratrix, 2004) are still tedious to handle. Some SMIL tools provide support for the "switch" element to define presentation alternatives; a comfortable interface for editing the different alternatives for many different contexts, however, is not provided. Consequently, we have been working on the approach in which a multimedia document is authored for one general context and is then "automatically" enriched by the different presentation alternatives needed for the expected user contexts in which the document is to be viewed (Boll et al., 1999). However, this approach is reasonable only for a limited number of presentation alternatives and limited presentation complexity in general.

Approaches that *dynamically* create personalized content are typically found on the Web, for example, Amazon.com (Amazon, 1996-2004) or MyYahoo (Yahoo!, 2002). However, these systems remain text-centric and are not occupied with the complex composition of media data in time and space into real multimedia presentations. On the pathway to an automatic generation of *personalized* multimedia presentations, we primarily find research approaches that address personalized *media* presentations only: For example, the home-video editor Hyper-Hitchcock (Girgensohn et al., 2003; Girgensohn et al., 2001) provides a preprocessing of a video such that users can interactively select clips to create their personal video summary. Other approaches create summaries of music or video (Kopf et al., 2004; Agnihotri et al., 2003). However, the systems provide an intelligent and intuitive access to large sets of (continuous) media rather than a dynamic creation of individualized content. An approach that addresses personalization for videos can be found, for example, with IBM's Video Semantic Summarization System (IBM Corporation, 2004a) which is, however, still concentrating on one single media type.

Towards *personalized multi*media we find interesting work in the area of *adaptive hypermedia* systems which has been going on for quite some years now (Brusilovsky 1996; Wu et al., 2001; De Bra et al., 1999a, 2000, 2002b; De Carolis et al., 1998, 1999). The adaptive hypermedia system AHA! (De Bra et al., 1999b, 2002a, 2003) is a prominent example here which also addresses the authoring aspect (Stash & De Bra, 2003), for example, in adaptive educational hypermedia applications (Stash et al., 2004). However, though these and further approaches integrate media elements in their adaptive hypermedia presentations, synchronized multimedia presentations are not in their focus.

Personalized or adaptive *user interfaces* allow the navigation and access of information and services in a customized or personalized fashion. For example, work done in the area of personalized agents and avatars considers "presentation generation" exploiting natural language generation and visual media elements to animate the agents and avatars (de Rosis et al., 1999). These approaches address the human computer interface; the general issue of dynamically creating arbitrary personalized multimedia content that meets the user's information needs is not in their research focus.

A very early approach towards the dynamic creation of multimedia content is the Coordinated Multimedia Explanation Testbed (COMET), which is based on an expert-system and different knowledge databases and uses constraints and plans to actually generate the multimedia presentations (Elhadad et al., 1991; McKeown et al., 1993). Another interesting approach to automate the multimedia authoring process has been developed at the DFKI in Germany by the two knowledge-based systems, WIP (Knowledge-based Presentation of Information) and PPP (Personalized Plan-based Presenter). WIP is a knowledge-based presentation system that automatically generates instructions for the maintenance of technical devices by plan generation and constraint solving. PPP enhances this system by providing a lifelike character to present the multimedia content and by considering the temporal order in which a user processes a presentation (André, 1996; André & Rist, 1995,1996). Also a very interesting research approach towards the dynamic generation of multimedia presentations is the Cuypers system (van Ossenbruggen et al., 2000) developed at the CWI. This system employs constraints for the description of the intended multimedia programming and logic programming for the generation of a multimedia document (CWI, 2004). The multimedia document group at INRIA in France developed within the Opéra project a generic architecture for the automated construction of multimedia presentations based on transformation sheets and constraints (Villard, 2001). This work is continued within the succeeding project Web, Accessibility, and Multimedia (WAM) with the focus on a negotiation and adaptation architecture for multimedia services for mobile devices (Lemlouma & Layaïda, 2003, 2004).

However, we find limitations with existing systems when it comes to their expressiveness and flexible personalized content creation support. Many approaches for personalization are targeted at a specific application domain in which they provide a very specific content personalization task. The existing research solutions typically use a declarative description like rules, constraints, style sheets, configuration files, and the like to express the dynamic, personalized multimedia content creation. However, they can solve only those presentation generation problems that can be covered by such a declarative approach; whenever a complex and application-specific personalization generation task is required, the systems find their limit and need additional programming to solve the problem. Additionally, the approaches we find usually rely on fixed data models for describing user profiles, structural presentation constraints, technical infrastructure, rhetorical structure, and so forth, and use these data models as an input to their personalization engine. The latter evaluates the input data, retrieves the most suitable content, and tries to most intelligently compose the media into a coherent aesthetic multimedia presentation. A change of the input data models as well as an adaptation of the presentation generator to more complex presentation generation tasks is difficult if not unfeasible. Additionally, for these approaches the border between the declarative descriptions for describing content personalization constraints and the additional programming needed is not clear and differs from solution to solution. This leads us to the development of a software framework that supports the development of personalized multimedia applications.

MULTIMEDIA PERSONALIZATION FRAMEWORK

Most of the research approaches presented above apply to text-centered information only, are limited with regard to the "personalizability", or are targeted at very specific application domains. As mentioned above, we find that existing research solutions in the field of multimedia content personalization provide interesting solutions. They typically use a declarative description like style sheets, transformation rules, presentation constraints, configuration files, and the like to express the dynamic, personalized multimedia content creation. However, they can solve only those presentation generation problems that can be covered by such a declarative approach; whenever a complex and application-specific personalization generation task is required, the systems find their limit and need additional programming to solve the problem. To provide application developers with a general, domain independent support for the creation of personalized multimedia content we pursue a software engineering approach: the MM4U framework. With this framework, we propose a component-based object-oriented software framework that relieves application developers from general tasks in the context of multimedia content personalization and lets them concentrate on the application domain-specific tasks. It supports the dynamic generation of arbitrary personalized multimedia presentations and therewith provides substantial support for the development of personalized multimedia applications. The framework does not reinvent multimedia content creation but incorporates existing research in the field and also can be extended by domain and application-specific solutions. In the following subsection we identify by an extensive study of related work and our own experiences the general design goals of this framework. In the next subsection, we present the general design of the MM4U framework, and then we present a detailed insight into the framework's layered architecture in the last subsection.

General Design Goals for the MM4U Framework

The overall goal of MM4U is to simplify and to reduce the costs of the development process of personalized multimedia applications. Therefore, the MM4U framework has to provide the developers with support for the different tasks of the multimedia personalization process as shown in Figure 2. These tasks comprise assistance for the access to media data and associated meta data as well as user profile information and the technical characteristics of the end device. The framework must also provide for the selection and composition of media elements into a coherent multimedia presentation. Finally, the personalized multimedia content must be created for delivery and rendering on the user's end device.

In regard to these different tasks, we conducted an extensive study of related work: In the area of user profile modeling we considered among others Composite Capability/Preference Profile (Klyne et al., 2003), FIPA Device Ontology Specification (Foundation for Intelligent Physical Agents, 2002), User Agent Profile (Open Mobile Alliance, 2003), Customer Profile Exchange (Bohrer & Holland, 2004), (Fink et al., 1997), and (Chen & Kotz, 2000). In regard to meta data modeling, we studied different approaches of modeling meta data and approaches for meta data standards for multimedia, for example, Dublin Core (Dublin Core Metadata Initiative, 1995-2003) and Dublin Core Extensions for Multimedia Objects (Hunter, 1999), Resource Description Framework (Beckett & McBride, 2003), and the MPEG-7 Multimedia content description standard (ISO/IEC JTC 1/SC 29/WG 11, 1999, 2001a-e). For multimedia composition we analyzed the features of multimedia document models, including SMIL (Ayars et al., 2001), SVG (Andersson et al., 2004b), Macrome-

dia Flash (Macromedia, 2004), Madeus (Jourdan et al., 1998), and ZYX (Boll & Klas, 2001). For the presentation of multimedia content, respective multimedia presentation frameworks were regarded including Java Media Framework (Sun Microsystems, 2004), MET++ (Ackermann 1996), and PREMO (Duke et al., 1999). Furthermore, other existing systems and general approaches for creating personalized multimedia content that were considered including the Cuypers engine (van Ossenbruggen et al., 2000) and the Standard Reference Model for Intelligent Multimedia Presentation Systems (Bordegoni et al., 1997).

We also derived design requirements to the framework from first prototypes of personalized multimedia applications we developed in different fields such as a personalized sightseeing tour through Vienna (Boll, 2003), a personalized mobile paper chase game (Boll et al., 2003), and a personalized multimedia music newsletter.

From the extensive study of related work and the first experiences and requirements we gained from our prototypical applications, we developed the single layers of the framework. We also derived three general design goals for MM4U. These design goals are

- The framework is to be designed such that it is independent of any special application domain, that is, it can be used to generate arbitrary personalized multimedia content. Therefore, it provides general multimedia composition and personalization functionality and is flexible enough to be adapted and extended concerning the particular requirements of the concrete personalization functionalities a personalized application needs.
- The access to user profile information and media data with its associated meta data must be independent of the particular solutions for storage, retrieval, and processing of such data. Rather the framework should provide a unified interface for the access to existing solutions. With distinct interfaces for the access to user profile information and media data with associated meta data, it is the framework's task to use and exploit existing (research) profile and media storage systems for the personalized multimedia content creation.
- The third design goal for the framework is what we call presentation independence. The framework is to be independent of, for example, the technical characteristics of the end devices, their network connection, and the different multimedia output formats that are available. This means, that the framework can be used to generate equivalent multimedia content for the different users and output channels and their individual characteristics. This *multichannel usage* implies that the personalized multimedia content generation task is to be partitioned into a composition of the multimedia content in an internal *representation format* and its later transformation into arbitrary (preferably standardized) *presentation formats* that can be rendered and displayed by end devices.

These general design goals have a crucial impact on the structure of the multimedia personalization framework, which we present in the following section.

General Design of the MM4U Framework

A software framework like MM4U is a semifinished software architecture, providing a software system as a generic application for a specific domain (Pree, 1995). The MM4U framework comprises components, which are bound together by their interaction (Szyperski et al., 2002), and realizes generic support for personalized multimedia applications. Each component is realized as an

object-oriented framework and consists of a set of abstract and concrete classes. Depending on the usage of a framework, the so-called "white-box" and "black-box" frameworks can be distinguished (respectively, white-box and gray-box reuse). A framework is used as a black-box if the concrete application that uses the framework adapts its functionality by different compositions of the framework's classes. In this case the concrete application uses only the built-in functionality of the framework, that is, those modules with which the framework is already equipped. In contrast, the functionality of a white-box framework is refined or extended by a concrete application, by adding additional modules through inheritance of (abstract) classes. Between these two contrasts arbitrary "shades of gray" are possible (Szyperski et al., 2002). The design of the MM4U framework lies somewhere in the middle between pure black-box and pure white-box. Being a domain independent framework, MM4U needs to be configured and extended to meet the specific requirements of a concrete personalized multimedia application. The framework provides many modules, for example, to access media data and associated meta data, user profile information, and generates the personalized multimedia content in a standardized output format that can be reused for different application areas (black-box usage). For the very application-specific personalization functionality, the framework can be extended correspondingly (white-box usage).

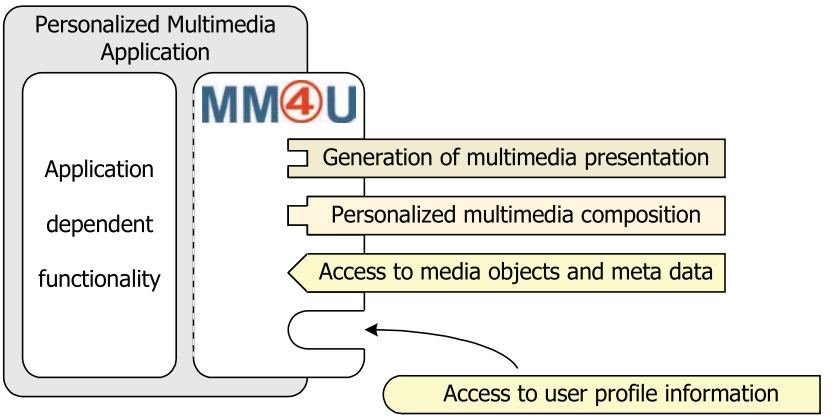
The usage of the MM4U framework by a concrete personalized multimedia application is illustrated schematically in Figure 3. The personalized multimedia application uses the functionality of the framework to create personalized multimedia content, and integrates it in whatever application dependent functionality is needed, either by using the already built-in functionality of the framework or by extending it for the specific requirements of the concrete personalized multimedia application.

With respect to the multimedia software development process the MM4U framework assists the computer scientists during the design and implementation phase. It alleviates the time-consuming multimedia content assembly task and lets the computer scientists concentrate on the development of the actual application. The MM4U framework provides functionality for the single tasks of the personalization engine as described in the section on Dynamic Authoring of Personalized Content. It offers the computer scientists support for integrating and accessing user profile information and media data, selecting media elements according to the user's profile information, composing these elements into coherent multimedia content, and generating this content in standardized multimedia document formats to be presented on the user's end device.

When designing a framework, the challenge is to identify the points where the framework should be flexible, that is, to identify the semantic aspects of the framework's application domain that have to be kept flexible. These points are the so-called *hot spots* and represent points or sockets of the intended flexibility of a framework (Pree, 1995). Each hot spot constitutes a well-defined interface where proper modules can be plugged in. When designing the MM4U framework we identified hot spots where adequate modules for supporting the personalization task can be plugged in that provide the required functionality.

As depicted in Figure 3, the MM4U framework provides four types of such hot spots, where different types of modules can be plugged in. Each hot spot represents a particular task of the personalization process. The hot spots can be realized by plugging in a module that implements the hot spot's functionality for a concrete personalized multimedia application. These modules can be both application-dependent and application-independent. For example, the access to media data and associated meta data is not necessarily application-dependent, whereas the composition

Figure 3. Usage of the MM4U framework by a personalized multimedia application



of personalized multimedia content can be heavily dependent on the concrete application.

After the general design of the framework, we take a closer look at the concrete architecture of MM4U and its components in the next section.

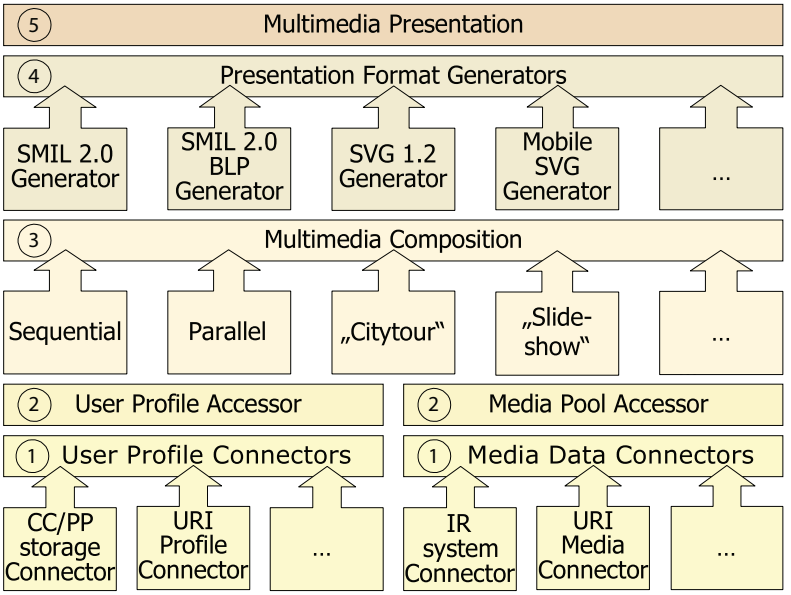
Design of the Framework Layers

For supporting the different tasks of the multimedia personalization process, which are the access to user profile information and media data, selection and composition of media elements into a coherent presentation, rendering and display of the multimedia presentation on the end device, a layered architecture seems to be best suited for MM4U. The layered design of the framework is illustrated in Figure 4. Each layer provides modular support for the different tasks of the multimedia personalization process. The access to user profile information and media data are realized by the layers (1) and (2), followed by the two layers (3) and (4) in the middle for composition of the multimedia presentation in an internal object-oriented representation and its later transformation into a concrete presentation output format. Finally, the top layer (5) realizes the rendering and display of the multimedia presentation on the end device.

To be most flexible for the different requirements of the concrete personalized multimedia applications, the framework's layers allow extending the functionality of MM4U by embedding additional modules as indicated by the empty boxes with dots. In the following descriptions, the features of the framework are described along its different layers. We start from the bottom of the architecture and end with the top layer.

(1) *Connectors*: The User Profile Connectors and the Media Data Connectors bring the user profile data and media data into the framework. They integrate existing systems for user profile stores, media storage, and retrieval solutions. As there are many different systems and formats available for user profile information, the User Profile Connectors abstract from the actual access to and retrieval of user profile information and provide a unified interface to the profile information. With this component, the different formats and structures of user profile models can be made accessible via a unified interface. For example, a flexible URIProfileConnector we developed for our demonstrator applications gains access to

Figure 4. Overview of the multimedia personalization framework MM4U



user profiles over the Internet. These user profiles are described as hierarchical ordered key-value pairs. This is a quite simple model but already powerful enough to allow effective pattern-matching queries on the user profiles (Chen & Kotz, 2000). However, as shown in Figure 4 also a User Profile Connector for the access to, for example, a Composite Capability/Preference Profile (CC/PP) server could be plugged into the framework.

On the same level, the Media Data Connectors abstract from the access to media elements in different media storage and retrieval solutions that are available today with a unified interface. The different systems for storage and content-based retrieval of media data are interfaced by this component. For example the URIMediaConnector, we developed for our demonstrator applications, provides a flexible access of media objects and its associated meta data from the Internet via http or ftp protocols. The meta data is stored in a single index file, describing not only the technical characteristics of the media elements and containing the location where to find the media elements in the Internet, but also comprise additional information about them, for example, a short description of what is shown in a picture or keywords for which one can search. By analogy with the access to user profile information, another Media Data Connector plugged into the framework could provide access to other media and meta data sources, for example, an image retrieval (IR) system like IBM's QBIC (IBM Corporation, 2004b).

The Media Data Connector supports the query of media elements by the client application (client-pull) as well as the automatic notification of the personalized application when a new media object arises in the media

database (server-push). The latter is required, for example, by the personalized multimedia sports news ticker (see the section about Sports4U) which is based on a multimedia event space (Boll & Westermann, 2003).

(2) *Accessors*: The User Profile Accessor and the Media Pool Accessor provide the internal data model of the user profiles and media data information within the system. Via this layer the user profile information and media data needed for the desired content personalization are accessible and processable for the application. The Connectors and Accessors are designed such that they are *not* reinventing existing systems for user modeling or multimedia content management. They, rather, provide a seamless integration of the systems by distinct interfaces and comprehensive data models. In addition, when a personalized multimedia application uses more than one user profile database or media database, the Accessor layer encapsulates the resources so that the access to them is transparent to the client application.

While the following layer (3) to (5) each constitute single components within the MM4U framework, the Accessor layer and Connectors layer do not. Instead the left side and the right side of the layers (1) and (2), i.e., the User Profile Accessor and User Profile Connectors as well as the Media Pool Accessor and Media Data Connectors, each form one component in MM4U.

(3) *Multimedia Composition*: The Multimedia Composition component comprises abstract operators in compliance with the composition capabilities of multimedia composition models like SMIL, Madeus, and ZYX, which provide complex multimedia composition functionality. It employs the data from the User Profile Accessor and the Media Pool Accessor for the multimedia composition task. The Multimedia Composition component is developed as such that it enables to develop additional, possibly more complex or application-specific composition operators that can be seamlessly plugged-in into the framework. Result of the multimedia composition is an internal object-oriented representation of the personalized multimedia content independent of the different presentation formats.

(4) *Presentation Format Generators*: The Presentation Format Generators work on the internal object-oriented data model provided by the Multimedia Composition component and convert it into a standardized presentation format that can be displayed by the corresponding multimedia player on the client device. In contrast to the multimedia composition operators, the Presentation Format Generators are completely independent of the concrete application domain and only rely on the targeted output format. In MM4U, we have already developed Presentation Format Generators for SMIL 2.0, the Basic Language Profile (BLP) of SMIL 2.0 for mobile devices (Ayars et al., 2001), SVG 1.2, Mobile SVG 1.2 (Andersson et al., 2004a) comprising SVG Tiny for multimedia-ready mobile phones and SVG Basic for pocket computers like Personal Digital Assistants (PDA) and Handheld Computers (HHC), and HTML (Raggett et al., 1998). We are currently working on Presentation Format Generators for Macromedia Flash (Macromedia, 2004) and other multimedia document model formats including HTML+TIME, the 3GPP SMIL Language Profile (3rd Generation Partnership Project, 2003b), which is a subset of SMIL used for scene description within the Multimedia Messaging Service (MMS) interchange format (3rd Generation Partnership Project,

2003a), and XMT-Omega, a high-level abstraction of MPEG-4 based on SMIL (Kim et al., 2000).

(5) *Multimedia Presentation*: The Multimedia Presentation component on top of the framework realizes the interface for applications to actually play the presentation of different multimedia presentation formats. The goal here is to integrate existing presentation components of the common multimedia presentation formats like SMIL, SVG, or HTML+TIME which the underlying Presentation Format Generator produces. So the developers benefit from the fact that only players for standardized multimedia formats need to be installed on the user's end device and that they must not spend any time and resources in developing their own render and display engine for their personalized multimedia application.

The layered architecture of MM4U permits easy adaption for the particular requirements that can occur in the development of personalized multimedia applications. So special user profile connectors as well as media database connectors can be embedded into the Connectors layer of the MM4U framework to integrate the most diverse and individual solutions for storage, retrieval and gathering for user profile information and media data. With the ability to extend the Multimedia Composition layer by complex and sophisticated composition operators, arbitrary personalization functionality can be added to the framework. The Presentation Format Generator component allows integrating any output format into the framework to support most different multimedia players that are available for the different end devices.

The personalized selection and composition of media elements and operators into a coherent multimedia presentation is the central task of the multimedia content creation process which we present in more detail in the following section.

CREATING PERSONALIZED MULTIMEDIA CONTENT

The MM4U framework provides the general functionality for the dynamic composition of media elements and composition operators into a coherent personalized multimedia presentation. Having presented the framework layers in the previous section, we now look in more detail how the layers contribute to the different tasks in the general personalization process as shown in Figure 2. The Media Data Accessor layer provides the personalized selection of media elements by their associated meta data and is described in the next subsection. The Multimedia Composition layer supports the composition of media elements into time and space in the internal multimedia representation format in three different manners, which are presented in detail in the the next three subsections, and the final subsection describes the last step, the transformation of the multimedia content in internal document model to an output format that is actually delivered to and rendered by the client devices. This is supported by the Presentation Format Generators layer.

Personalized Multimedia Content Selection

For creating personalized multimedia content, first those media elements have to be selected from the media databases that are most relevant to the user's request. This personalized media selection is realized by the Media Data Accessor and Media Data Connector component of the framework. For the actual personalized selection of media elements, a context object is created within the Multimedia Composition layer carrying the user profile information, technical characteristics of the end device, and further application-specific information. With this context object, the unified interface of the Media Data Accessor for querying media elements is called. The context object is handed over to the concrete Media Data Connector

of the connected media database. Within the Media Data Connector, the context object is mapped to the meta data associated with the media elements in the database and those media elements are determined that match the request, that is, the given context object at best. It is important to note that the Media Data Accessor and Media Data Connector layer integrate and embrace existing multimedia information systems and modern content-based multimedia retrieval solutions. This means that the retrieval of the "best match" can only be left to the underlying storage and management systems. The framework can only provide for comprehensive and lean interfaces to these systems. This can be our own URIMediaServer accessed by the URIMediaConnector but also other multimedia databases or (multi)media retrieval solutions. The result set of the query is handed back by the Accessor to the composition layer.

For example, the context object for our mobile tourist guide application carries information about user interests and preferences with respect to the sights of the city, the display size of the end device, and the location for the tourist guide. The Media Data Connector, realized in this case by our URIMediaConnector, processes this context object and returns images and videos from those sights in Oldenburg that both match the user's interests and preferences as well as the limited display size of the mobile device.

Based on the personalized selection of media elements the Multimedia Composition layer provides the assembly of these media elements on three different manners, the basic, complex and sophisticated composition of multimedia content, which is described in the following sections.

Basic Composition Functionality

With the basic composition functionality the MM4U framework provides the basic bricks for composing multimedia content. It forms the basis for assembling the selected media elements into personalized multimedia documents and provides the means for realizing the central aspects of multimedia document models, that is, the temporal model, the spatial layout, and the interaction possibilities of the multimedia presentation. The temporal model of the multimedia presentation is determined by the temporal relationships between the presentation's media elements formed by the composition operators. The spatial layout expresses the arrangement and style of the visual media elements in the multimedia presentation. Finally, with the interaction model the user interaction of the multimedia presentation is determined, in order to let the user choose between different paths of a presentation. For the temporal model, we selected an interval-based approach as found in Duda & Keramane (1995). The spatial layout is realized by a hierarchical model for media positioning (Boll & Klas, 2001). For interaction with the user navigational and decision interaction are supported, as can be found with SMIL (Ayars et al., 2001) and MHEG-5 (Echiffre et al., 1998; International Organisation for Standardization, 1996).

A basic composition operator or *basic operator* can be regarded as an atomic unit for multimedia composition, which cannot be further broken down. Basic operators are quite simple but applicable for any application area and therefore most flexible. Basic temporal operators realize the temporal model, and basic interaction operators realize the interaction possibilities of the multimedia presentation, as specified above. The two basic temporal operators *Sequential* and *Parallel*, for example, can be used to present media elements one after the other in a sequence respectively to present media elements parallel at the same time. With basic temporal operators and media elements, the temporal course of the presentation can be determined like a slideshow as depicted in Figure 5. The operators are represented by white rectangles and the media elements by gray ones. The relation between the media elements and the basic operators is shown by the edges beginning with a filled circle at an operator and ending with

Figure 5. Slideshow as an example of assembled multimedia content

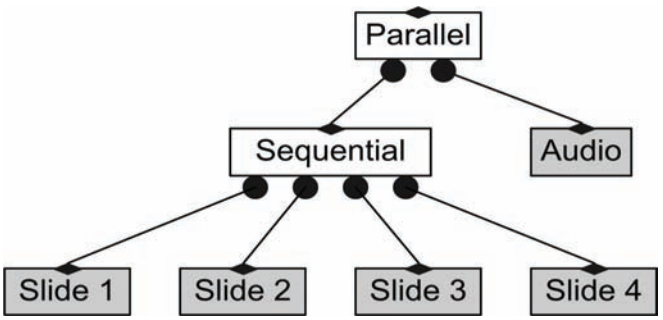
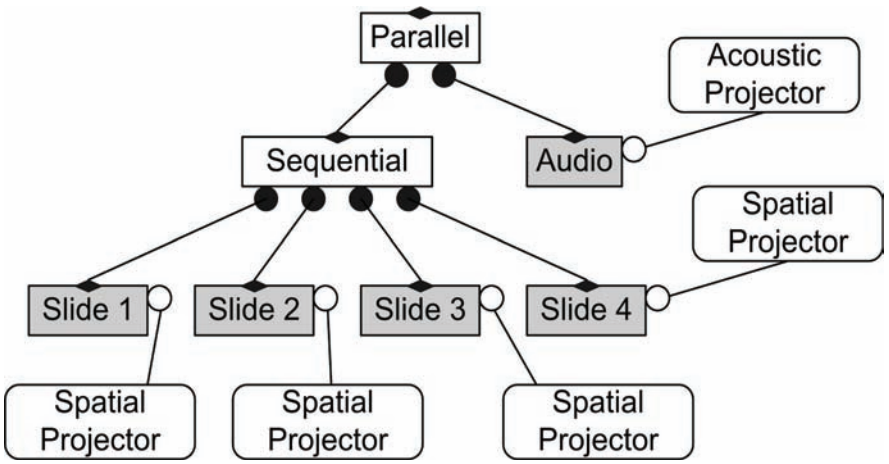


Figure 6. Adding layout to the slideshow example



a filled rhombus respectively a diamond at a media element or another operator. The semantics of the slideshow shown in Figure 5 are that it starts with the presentation of the root element, which is the Parallel operator. The semantics of the Parallel operator are that it shows the operators and media elements that are attached to it at the same time. This means that the audio file starts to play while simultaneously the Sequential operator is presented. The semantics of the Sequential operator are to show the attached media elements one after another, so while the audio file is played in the background, the four slides are presented in sequence.

Besides the basic composition *operators*, the so-called *projectors* are part of the Multimedia Composition layer. Projectors can be attached to operators and media elements to define, for example, the visual and acoustical layout of the multimedia presentation. Figure 6 shows the

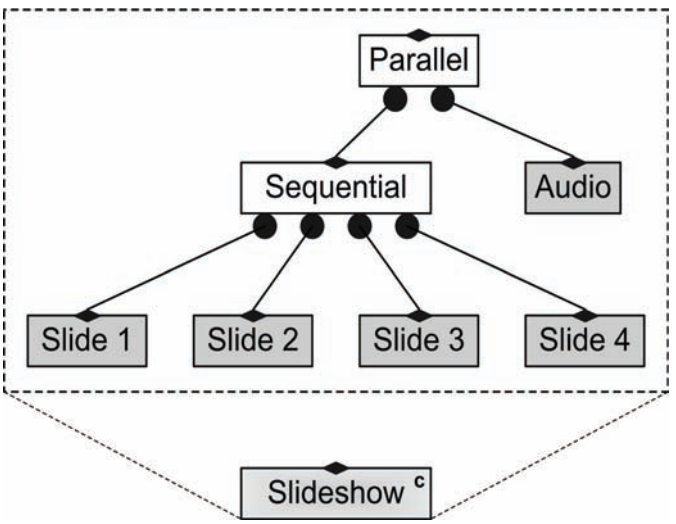
slideshow example from above with projectors attached. The spatial position as well as the width and height of the single slide media elements are determined by the corresponding *SpatialProjectors.* The volume, treble, bass, and balance of the audio medium is determined by the attached *AcousticProjector.*

Besides temporal operators, the Multimedia Composition component offers basic operators for specifying the interaction possibilities of the multimedia presentation. Interaction can be added, for example, by using the basic operator *InteractiveLink*. It defines a link represented by a single media element or a fragment of a multimedia presentation that is clickable by the user and a target presentation the user receives if he or she clicks on the link.

The description above presents some examples of the basic composition functionality the MM4U framework offers. The framework comprises customary composition operators for creating multimedia content as provided by modern multimedia presentation formats like SMIL and SVG. Even though the basic composition functionality does not reflect the fancy features of some of today's multimedia presentation formats, it supports the very central multimedia features of modeling time, space, and interaction. This allows the transformation of the internal document model into many different multimedia presentation formats for different end devices.

With the basic multimedia composition operators the framework offers, arbitrary multimedia presentations can be assembled. However, so far the MM4U framework provides "just" basic multimedia composition functionality. In the same way that one would use an authoring tool to create SMIL presentations, for example, the GR*i*NS editor (Oratrix, 2004), one can also use a corresponding authoring tool for the basic composition operators the MM4U framework offers to create multimedia content. For reasons of reusing parts of the created multimedia presentations, for example, a menu bar or a presentation's layout, and for convenience, there is a need for more complex and application-specific composition operators that provide a more convenient support for creating the multimedia content.

Figure 7. The slideshow examples as a complex composition operator

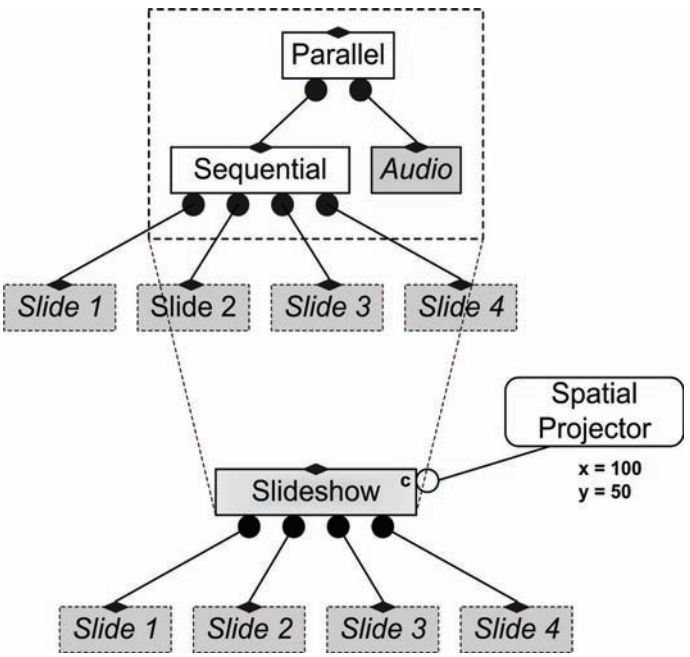


Complex Composition Functionality

For creating presentations that are more complex, the Multimedia Composition layer provides the ability to abstract from basic to complex operators. A complex composition operator encapsulates the composition functionality of an arbitrary number of basic operators and projectors and provides the developers with a more complex and application-specific building block for creating the multimedia content. Complex composition operators are composed of basic and others complex operators. As complex composition operators not only embed basic but also other complex operators, they provide for a reuse of composition operators. In contrast to the basic operators, the complex composition operators can be dismantled into their individual parts. Figure 7 depicts a complex composition operator for our slideshow example. It encapsulates the media elements, operators, and projectors of the slideshow (the latter are omitted in the diagram to reduce complexity). The complex operator *Slideshow*, indicated by a small "c" symbol in the upper right corner, represents an encapsulation of the former slideshow presentation in a complex object and forms itself a building block for more complex multimedia composition.

Complex operators, as described above, define fixed encapsulated presentations. Their temporal flow, spatial layout, and the used media elements cannot be changed subsequently. However, a complex composition operator does not necessarily need to specify all media elements, operators, and projectors of the respective multimedia document tree. Instead, to be more flexible, some parts can be intentionally left open. These parts constitute the parameters of a complex composition operator and have to be filled in for concrete usage of these operators. Such parameterized complex composition operators are one means to define *multimedia composition templates* within the MM4U framework. However, only prestructured multimedia content can be created with

Figure 8. The slideshow example as parameterized complex composition operator



these templates, since the complex composition operators can only encapsulate presentations of a fixed structure.

Figure 8 shows the slideshow example as a parameterized complex composition operator. In this case, the complex operator Slideshow comprises the two basic operators Parallel and Sequential. The Slideshow's parameters are the place holders for the single slides and have to be instantiated when the operator is used within the multimedia composition. The slideshow's audio file is already preselected. In addition, the parameters of a complex composition operator can be typed, that is, they expect a special type of operator or media element. The Slideshow operator would expect visual media elements for the parameters *Slide 1* to *Slide 4*. To indicate the complex operator's parameters, they are visualized by rectangles with dotted lines. The preselected audio file is already encapsulated in the complex operator as illustrated in Figure 7.

In the same way projectors are attached to basic operators in the basic composition functionality section, they can also be attached to complex operators. The SpatialProjector attached to the Slideshow operator shown in Figure 8 determines that the slideshow's position within a multimedia presentation is the position x = 100 pixel and y = 50 pixel in relation to the position of its parent node.

With basic and complex composition operators one can build multimedia composition functionality that is equivalent to the composition functionality of advanced multimedia document models like Madeus (Jourdan et al., 1998) and ZYX (Boll & Klas, 2001). Though complex composition operators can have an arbitrary number of parameters and can be configured individually each time they are used, the internal structure of complex operators is still static. Once a complex operator is defined, the number of parameters and their type are fixed and cannot be changed. Using a complex composition operator can be regarded as filling in fixed composition templates with suitable media elements. Personalization can only take place in selecting those media elements that fit the user profile information at best. For the dynamic creation of *personalized* multimedia content even more sophisticated composition functionality is needed, that allows the composition operators to change the structure of the generated multimedia content at runtime. To realize such sophisticated composition functionality, additional composition logic needs to be included into the composition operators, which cannot be expressed anymore even by the mentioned advanced document models we find in the field.

Sophisticated Composition Functionality

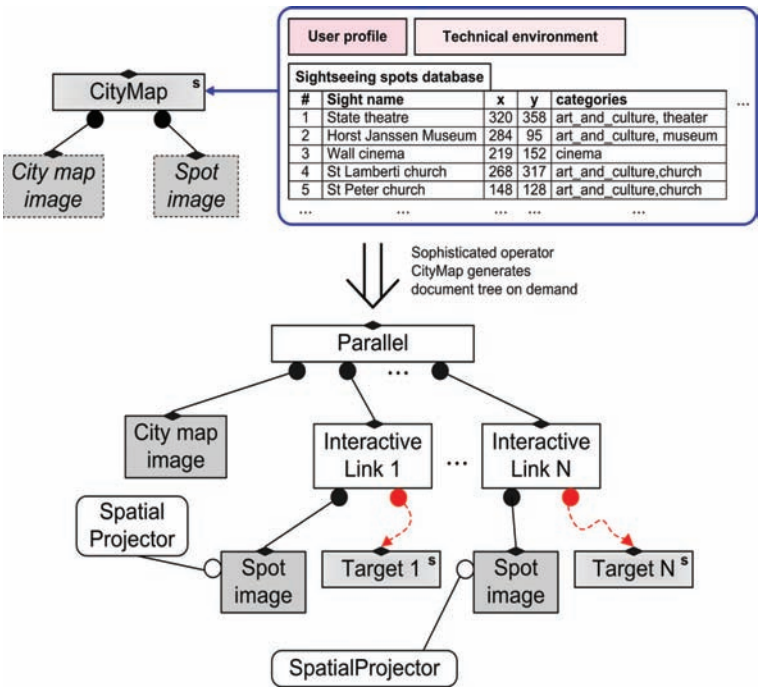
With basic and complex composition functionality, we already provide the dynamic composition of prestructured multimedia content by parameterized multimedia composition templates. However, such templates are only flexible concerning the selection of the concrete composition parameters. To achieve an even more flexible dynamic content composition, the framework provides sophisticated composition operators, which allow determining the document structure and layout during creation time by additional composition logic. Multimedia composition templates defined by using such sophisticated composition operators are no longer limited to creating prestructured multimedia content only, but determine the document structure and layout of the multimedia content on-the-fly and, depending on the user profile information, the characteristics of the used end device, and any other additional information. The latter can be, for example, a database containing sightseeing information. Such sophisticated composition operators exploit the basic and complex composition operators the MM4U framework offers but allow more flexible, possibly application-specific multimedia composition and personalization functionality with their additional composition logic. This compo-

sition logic can be realized by using document structures, templates, constraints and rules, or by plain programming. Independent of how the sophisticated multimedia content composition functionality is actually realized, the result of this composition process is always a multimedia document tree that consists of basic and complex operators, projectors, as well as media elements. In our graphical notation, sophisticated composition operators are represented in the same way as complex operators, but are labeled with a small "s" symbol in the upper right corner.

Figure 9 shows an example of a parameterized sophisticated composition operator, the *CityMap*. This operator provides the generation of a multimedia presentation containing a city map image together with a set of available sightseeing spots on it. The parameters of this sophisticated operator are the city map image of arbitrary size and a spot image used for presenting the sights on the map. Furthermore, the CityMap operator reads out the positions of the sights on a reference map (indicated by the table on the right) and automatically recalculates the positions in dependence of the size of the actual city map image. Which spots are selected and actually presented on the city map depends on the user profile, in particular the types of sights he or she is interested in, and the categories a sight belongs to. In addition, the size of the city map image is selected to best fit the display of the end device. The CityMap operator is used within our personalized city guide prototype presented in the section on Sightseeing 4U and serves there for Desktop PCs as well as mobile devices.

The multimedia document tree generated by the CityMap operator is shown in the bottom part of Figure 9. Its root element constitutes the Parallel operator. Attached to it are the image of the city map and a set of InteractiveLink operators. Each InteractiveLink represents a spot on the city map, instantiated by the spot image. The

Figure 9. Insight into the sophisiticated composition operator CityMap



#	Sight name	x	y	categories
1	State theatre	320	358	art_and_culture, theater
2	Horst Janssen Museum	284	95	art_and_culture, museum
3	Wall cinema	219	152	cinema
4	St Lamberti church	268	317	art_and_culture,church
5	St Peter church	148	128	art_and_culture,church

user can click on the spots to receive multimedia presentations with further information about the sights. The positions of the spot images on the city map are determined by the SpatialProjectors. The personalized multimedia presentations about the sights are represented by the sophisticated operators Target 1 to Target N.

The CityMap operator is one example of extending the personalization functionality of the MM4U framework by a sophisticated application-specific multimedia composition operator, here in the area of (mobile) tourism applications. This operator, for example, is developed by programming the required dynamic multimedia composition functionality. However, the realization of the internal composition logic of sophisticated operators is independent of the used technology and programming language. The same composition logic could also be realized by using a different technology, for example, a constraint-based approach. Though the actual realization of the personalized multimedia composition functionality would be different, the multimedia document tree generated by this rule-based sophisticated operator would be the same as depicted in Figure 9.

Sophisticated composition operators allow embracing the most different solutions for realizing personalized multimedia composition functionality. This can be plain programming as in the example of the CityMap operator, document structures and templates that are dynamically selected according to the user profile information and filled in with media elements relevant to the user, or systems describing the multimedia content generation by constraints and rules.

The core MM4U framework might not offer all kinds of personalized multimedia composition functionality one might require, since the personalization functionality always depends on the actual application to be developed and thus can be very specific. Instead, the framework provides the basis to develop sophisticated personalized multimedia composition operators, such that every application can integrate its own personalization functionality into the framework. So, every sophisticated composition operator can be seen as a "small application" itself that can conduct a particular multimedia personalization functionality. This small application can be reused within others and thus extends the functionality of the framework. The Multimedia Composition component allows a seamless plug-in of arbitrary sophisticated composition operators into the MM4U framework. This enables the most complex personalized multimedia composition task to be "just" plugged into the system and to be used by a concrete personalized multimedia application.

With the sophisticated composition operators, the MM4U framework provides its most powerful and flexible functionality to generate arbitrary personalized multimedia content. However, this multimedia content is still represented in an internal document model and has to be transformed into a presentation format that can be rendered and displayed by multimedia players on the end devices.

From Multimedia Composition to Presentation

In the last step of the personalization process, the personalized multimedia content represented in our internal document model is transformed by the Presentation Format Generators component into one of the supported standardized multimedia presentation formats, which can be rendered and displayed on the client device. The output format of the multimedia presentation is selected according to the user's preferences and the capabilities of the end device, that is, the available multimedia players and the multimedia presentation formats they support.

The Presentation Format Generators adapt the characteristics and facilities of the internal document model provided by the Multimedia Composition layer in regard of the used time model, spatial layout, and interaction possibilities

to the particular characteristics and syntax of the concrete presentation format. For example, the spatial layout of our internal document model is realized by a hierarchical model that supports the positioning of media elements in relation to other media elements. This relative positioning is supported by most of today's presentation formats, for example, SMIL, SVG, and XMT-Omega. However, there exist multimedia presentation formats that do not support such a hierarchical model and only allow an absolute positioning of visual media elements in regard to the presentation's origin as, for example, the Basic Language Profile of SMIL, 3GPP SMIL, and Macromedia's Flash. In this case, the Presentation Format Generators component transforms the hierarchically organized spatial layout of the internal document model to a spatial layout of absolute positioning. How the transformation of the spatial model is actually performed and how the temporal model and interaction possibilities of the internal document model are transformed into the characteristics and syntax of the concrete presentation formats is intentionally omitted in this book chapter due to its focus on the composition and assembly of the personalized multimedia content and is described in Scherp and Boll (2005).

IMPACT OF PERSONALIZATION TO THE DEVELOPMENT OF MULTIMEDIA APPLICATIONS

The multimedia personalization framework MM4U presented so far provides support to develop sophisticated personalized multimedia applications. Involved parties in the development of such applications are typically a heterogeneous team of developers from different fields including media designers, computer scientists, and domain experts. In this section, we describe what challenges *personalization* brings to the development of personalized multimedia applications and how and where the MM4U framework can support the developer team to accomplish their job.

In the next subsection, the general software engineering issues in regard to personalization are discussed. We describe how personalization affects the single members of the heterogeneous developer team and how the MM4U framework supports the development of personalized multimedia applications. The challenges that arise with creating personalized multimedia content by the domain experts using an authoring tool are presented in the following subsection. We also introduce how the MM4U framework can be used to develop a domain-specific authoring tool in the field of e-learning content, which aims to hide the technical details of content authoring from the authors and lets them concentrate on the actual creation of the personalized multimedia content.

Influence of Personalization to Multimedia Software Engineering

We observe that software engineering support for multimedia applications such as proper process models and development methodologies are not likely to be found in this area. Furthermore, the existing process models and development methodologies for multimedia applications as for example Rout and Sherwood (1999) and Engels et al. (2003) do not support personalization aspects. However, personalization requirements complicate the software development process even more and increase the development costs, since every individual alternative and variant has to be anticipated, considered, and actually implemented. Therefore, there is a high demand in supporting the development process of such applications. In the following paragraphs, we first introduce how personalization affects the software development process with respect to the multimedia content creation process in general. Then we identify what support the developers of

personalized multimedia applications need and consider where the MM4U framework supports the development process.

Since the term *personalization* profoundly depends on the application's context, its meaning has ever to be reconsidered when developing a personalized application for a new domain. Rossi et al. (2001) claim that personalization should be considered directly from the beginning when a project is conceived. Therefore, the first activity when developing a personalized multimedia application is to determine the personalization requirements, that is, which aspects of personalization should be supported by the actual application. For example, in the case of an e-learning application the personalization aspects consider the automatic adaptation to the different learning styles of the students and their prior knowledge about the topic. In addition, different degrees of difficulty should be supported by a personalized e-learning application. In the case of a personalized mobile tourism application, however, the user's location and his or her surroundings would be of interest for personalization instead. These personalization aspects must be kept in mind during every activity throughout the whole development process. The decision regarding which personalization aspects are to be supported has to be incorporated in the analysis and design of the personalized application and will hopefully entail a flexible and extendible software design. However, this increases the overall complexity of the application to be developed and automatically leads to a higher development effort including longer development duration and higher costs. Therefore, a good requirement analysis is crucial when developing personalized applications lest one dissipates one's energies in bad software design with respect to the personalization aspects.

When transferring the requirements for developing personalized software to the specific requirements of personalized multimedia applications one can say that it affects all members of the developer team: the domain expert, the media designers, and the computer scientists, and putting higher requirements to them.

The domain expert normally contributes to the development of multimedia applications by providing input to draw storyboards of the specific application's domain. These storyboards are normally drawn by media designers and are the most important means to communicate the later application's functionality within the developer team. When personalization comes into account, it is difficult to draw such storyboards, because of the many possible alternatives and different paths in the application that are implicated with personalization. Consequently, the storyboards change in regard to, for example, the individual user profiles and the end devices that are used. When drawing storyboards for a *personalized* multimedia application, those points in the storyboard have to be identified and visualized where personalization is required and needed. Storyboards have to be drawn for every typical personalization scenario concerning the concrete application. This drawing task should be supported by interactive graphical tools to create "personalized" storyboards and to identify reusable parts and modules of the content.

It is the task of the media designer in the development of multimedia applications to plan, acquire, and create media elements. With personalization, media designers have to think additionally about the usage of media elements for personalization purposes, that is, the media elements have to be created and prepared for different contexts. When acquiring media elements, the media designers must consider for which user context the media elements are created and what aspects of personalization are to be supported, for example, different styles, colours, and spatial dimensions. Possibly a set of quite similar media assets have to be developed, that only differ in certain aspects. For example, an image or video has to be transformed for different end device resolutions, colour depth, and network connections. Since personalization means to (re)assemble

existing media elements into a new multimedia presentation, the media designers will also have to identify reusable media elements. This means that additionally the storyboards must already capture the personalization aspects. Not only the content but also the layout of the multimedia application can change depending on the user context. So, the media designers have to create different visual layouts for the same application to serve the needs of different user groups. For example, an e-learning system for children would generate colourful multimedia presentations with many auditory elements and a comic-like virtual assistant, whereas the system would present the same content in a much more factual style for adults. This short discussion shows that personalization already affects the storyboarding and media acquisition. Creating media elements for personalized multimedia applications requires a better and elaborate planning of the multimedia production. Therefore, a good media production strategy is crucial, due to the high costs involved with the media production process. Consequently, the domain experts and the media designers need to be supported by appropriate tools for planning, acquiring, and editing media elements for personalized multimedia applications.

The computer scientists actually have to develop the multimedia personalization functionality of the concrete application. What this personalization functionality is depends heavily on the concrete application domain and is communicated with the domain experts and media designers by using personalized storyboards. With personalization, the design of the application has to be more flexible and more abstract to meet the requirements of changing user profile information and different end device characteristics. This is where the MM4U framework comes into play. It provides the computer scientists the general architecture of the personalized multimedia application and supports them in designing and implementing the concrete multimedia personalization functionality. When using the MM4U framework, the computer scientists must know how to use and to extend it. The framework provides the basis for developing both basic and sophisticated multimedia personalization functionality, as for example the Slideshow or the CityMap operator presented in the section on content. To assist the computer scientists methodically we are currently working on guidelines and checklists of how to develop the personalized multimedia composition operators and how to apply them. Consequently, the development of personalized multimedia applications by using the MM4U framework basically means to the computer scientists the design, development, and deployment of multimedia composition operators for generating personalized content. The concept of the multimedia personalization operators as introduced in the content section, that every concrete personalized multimedia application is itself a new composition operator increases reusage of existing personalization functionality. Furthermore, the interface design of the sophisticated operators makes it possible to embrace existing approaches that are able to generate multimedia document trees, for example, so it can be generated with the basic and complex composition functionality of the MM4U framework.

Influence of Personalization to Multimedia Content Authoring

Authoring of multimedia content is the process in which the multimedia presentations are actually created. This creation process is typically supported by graphical authoring tools, for example, Macromedia's Authorware and Director (Macromedia, 2004), Toolbook (Click2learn, 2001-2002), (Arndt, 1999), and (Gaggi & Celentano, 2002). For creating the multimedia content, the authoring tools follow different design philosophies and metaphors, respectively. These metaphors can be roughly categorized into script-based, card/page-based, icon-based, timeline-based, and object-based authoring (Rabin & Burns, 1996).

All these different metaphors have the same goal, to support authors in creating their content. Even though based on these metaphors a set of valuable authoring tools has been developed, these metaphors do not necessarily provide a suitable means for authoring personalized content.

From the context of our research project Cardio-OP we derived early experiences with personalized content authoring for domain experts in the field of cardiac surgery (Klas et al., 1999; Greiner & Rose, 1998; Boll et al., 2001). One of the tools developed by a project partner, the Cardio-OP Authoring Wizard, is a page-based easy-to-use multimedia authoring environment, enabling medical experts to compose a multimedia book on operative techniques in the domain of cardiac surgery for three different target groups, medical doctors, nurses, and students. The Authoring Wizard guides the author through the particular authoring steps and offers dialogues specifically tailored to the needs of each step. Coupled tightly with an underlying media server, the authoring wizard allows use of every precious piece of media data available at the media server in all of the instructional applications at different educational levels. This promotes reuse of expensively produced content in a variety of different contexts.

Personalization of the e-learning content is required here, since the three target groups have different views and knowledge about the domain of cardiac surgery. Therefore, the target groups require different information from such a multimedia book, presented on an adequate level of difficulty for each group.

However, the experiences we gained from deploying this tool show that it is hard to provide the domain authors with an adequate intuitive user interface for the creation of personalized multimedia e-learning content for three educational levels. It was a specific challenge for the computer scientists involved in the project to provide both media creation tools and multimedia authoring wizard that allow the domain experts to insert knowledge into the system, while at the same time hiding the technical details from them as much as possible.

On the basis of the MM4U framework, we are currently developing a "smart authoring tool" aimed for domain experts to create personalized multimedia e-learning content. The tool we are developing works at the what-you-see-is-what-you-get (WYSIWYG) level and can be seen as a specialized application employing the framework to create personalized content. The content source from which this personalized e-learning content is created constitutes the LEBONED repositories. Within the LEBONED project (Oldenettel & Malachinski, 2003) digital libraries are integrated into learning management systems. Using the content managed by the LEBONED system for new e-learning units, a multimedia authoring support is needed for assembling existing e-learning modules into new, possibly more complex, units. In the e-learning context, the background of the learners is very relevant for the content that meets the users learning demands — that means a personalized multimedia content can meet the user's background knowledge and interest much better than a one-size-fits-all e-learning unit. The creation of an e-learning unit on the other side cannot be supported by a mere automatic process. Rather the domain experts would like to control the assembly of the content because they are responsible for the content conveyed. The smart authoring tool guides the domain experts through the composition process and supports them in creating presentations that still provide flexibility to the targeted user context. In the e-learning context we can expect domain experts such as lecturers that want to create a new e-learning unit but do not want to be bothered with the technical details of (multimedia) authoring.

We use the MM4U framework to build the multimedia composition and personalization functionality of this smart authoring tool. For this, the Multimedia Composition component supports the creation and processing of arbitrary

document structures and templates. The authoring tool exploits this functionality for composition to achieve a document structure that is suitable just for that content domain and the targeted audience. The Media Data Accessor supports the authoring tool in those parts in which it lets the author choose from only those media elements that are suitable for the intended user contexts and that can be adapted to the user's infrastructure. Using the Presentation Format Generators, the authoring tool finally generates the presentations for the different end devices of the targeted users. Thus the authoring process is guided and specialized with regard to selecting and composing personalized multimedia content. For the development of this authoring tool, the framework fulfils the same function in the process of creating personalized multimedia content in a multimedia application as described in the previous section on the framework. However, the creation of personalized content is not achieved at once but step by step during the authoring process.

IMPLEMENTATION AND PROTOTYPICAL APPLICATIONS

The framework, its components, classes and interfaces, are specified using the Unified Modeling Language (UML) and has been implemented in Java. The development process for the framework is carried out as an iterative software development with stepwise refinement and enhancement of the framework's components. The redesign phases are triggered by the actual experience of implementing the framework but also by employing the framework in several application scenarios. In addition, we are planning to provide a beta version of the MM4U framework to other developers for testing the framework and to develop their own personalized multimedia applications with MM4U.

Currently, we are implementing several application scenarios to prove the applicability of MM4U in different application domains. These prototypes are the first stress test for the framework. At the same time the development of the sample applications gives us an important feedback about the comprehensiveness and the applicability of the framework. In the following sections, two of our prototypes that are based on the MM4U framework are introduced: In the Sightseeing4U subsection, a prototype of a personalized city guide is presented, and in the Sports4U subsection a prototype of a personalized multimedia sports news ticker is described.

Sightseeing4U: A Generic Personalized City Guide

Our first prototype using the MM4U framework is Sightseeing4U, a generic personalized city guide application (Scherp & Boll, 2004a, 2004b; Boll et al., 2004). It is applicable to develop personalized tourist guides for arbitrary cities, both for desktop PCs and mobile devices such as PDAs (Personal Digital Assistants). The generic Sightseeing4U application uses the MM4U framework and its modules as depicted in Figure 3. The concrete demonstrator we developed for our hometown Oldenburg in Northern Germany considers the pedestrian zone and comprises video and image material of about 50 sights. The demonstrator is developed for desktop PCs as well as PDAs (Scherp & Boll, 2004a). It supports personalization with respect to the user's interests, for example, churches, museums, and theatres, and preferences such as the favorite language. Depending on the specific sightseeing interests, the proper sights are automatically selected for the user. This is realized by category matching of the user's interests with the meta data associated to the sights. Figure 10 and Figure 11 show some screenshots of our city guide application in different output formats and on different end devices. The presentation in Figure 10 is targeted at a user interested in culture, whereas the presentation in Figure 11 is generated for a user who is hungry and searches for a good

Figure 10. Screenshots of the city guide application for a user interested in culture (presentation generated in SMIL 2.0 and SMIL 2.0 BLP format, respectively)

(a) RealOne Player (RealNetworks, 2003) on a Desktop PC

(b) PocketSMIL Player (INRIA, 2003) on a PDA

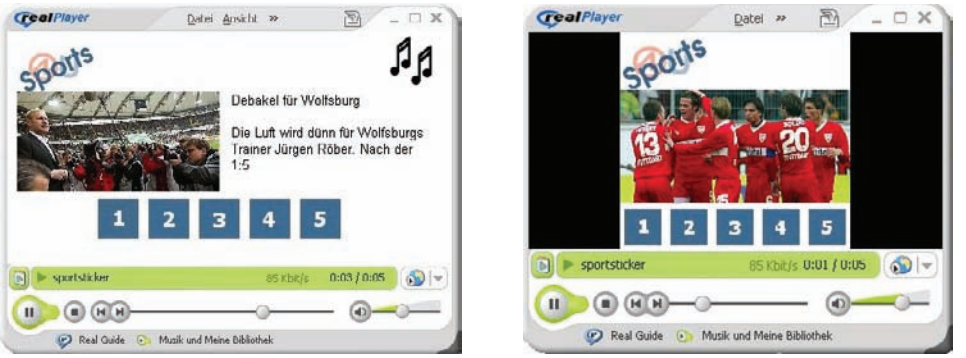
Figure 11. Screenshots of the Sightseeing4U prototype for a user searching for a good restaurant (output generated in SVG 1.2 and Mobile SVG format, respectively)

(a) Adobe SVG Plug-In (Adobe Systems, 2001) on a Tablet PC

(b) Pocket eSVG viewer (EXOR, 2001-2004) on a PDA

restaurant in Oldenburg. The different interests of the users result in different spots that are presented on the map of Oldenburg. When clicking on a certain spot the user receives a multimedia presentation with further information about the sight (see the little boxes where the arrows point). Thereby, the media elements for the multimedia presentation are automatically selected to fit the

Figure 12. Screenshots of the personalized sports newsticker Sport4U



end devices' characteristics best. For example, a user sitting at a desktop PC receives a high-quality video about the palace of Oldenburg as depicted in Figure 10a, while a mobile user gets a smaller video of less quality in Figure 10b. In the same way, the user searching for a good restaurant in Oldenburg receives either a high-quality video when using a Tablet PC as depicted in Figure 11a, or a smaller one that meets the limitations of the mobile device as shown in Figure 11b. If there is no video of a particular sight available at all, the personalized tourist guide automatically selects images instead and generates a slideshow for the user.

Sports4U: A Personalized Multimedia Sports News Ticker

A second prototype that uses our MM4U framework is the personalized multimedia sports news ticker called Sports4U. The Sports4U application exploits the MediÆther multimedia event space as introduced in (Boll & Westermann, 2003). The MediÆther is based on a decentralized peer-to-peer infrastructure and allows one to publish, to find, and to be notified about any kind of multimedia events of interest. In the case of Sports4U, the event space forms the media data basis of sports-related multimedia news events. A multimedia sports news event comprises data of different media types like describing text, a title, one or more images, an audio record, or a video clip. The personalized sports ticker application combines the multimedia data of the selected events, the available meta data, and additional information, for example, from a soccer player database. The application uses a sophisticated composition operator that automatically arranges these multimedia sports news to a coherent presentation. It regards possible constraints like running time limit and particular characteristics of the end device, like the limited display size of a mobile device. The result is a sports news presentation that can be, for example, viewed with an SMIL player over the Web as shown in Figure 12. With a suitable Media Data Connector the MediÆther is connected to the MM4U framework. This connector not only allows querying for media elements like the URIMediaConnector but also provides the notification of incoming multimedia events to the actual personalized application. Depending on the user context, the Sports4U prototype receives the sports news from the pool of sports events in the MediÆther that match the users profile. The Sports4U application alleviates the user from the time-consuming task of searching for sports news he or she might be interested in.

CONCLUSION

In this chapter, we presented an approach for supporting the creation of personalized multimedia content. We motivated the need of technology to handle the flood of multimedia information that allows for a much targeted, individual management and access to multimedia content. To give a better understanding of the content creation process we introduced the general approaches in multimedia data modeling and multimedia authoring as we find it today. We presented how the need for personalization of multimedia content heavily affects the multimedia content creation process and can only result in a dynamic, (semi)automatic support for the personalized assembly of multimedia content. We looked into existing related approaches ranging from personalization in the text-centric Web context over single media personalization to the personalization of multimedia content. Especially for complex personalization tasks we observe that an (additional) programming is needed and propose a software engineering support with our Multimedia for you Framework (MM4U).

We presented the MM4U framework concept in general and, in more detail, the single layers of the MM4U framework: access to user profile information, personalized media selection by meta data, composition of complex multimedia presentations, and generation of different output formats for different end devices. As a central part of the framework, we developed multimedia composition operators which create multimedia content in an internal model and representation for multimedia presentations, integrating the composition capabilities of advanced multimedia composition models. Based on this representation, the framework provides so-called generators to dynamically create different context-aware multimedia presentations in formats such as SMIL and SVG. The usage of the framework and its advantages has been presented in the context of multimedia application developers but also in the specific case of using the framework's specific features for the development of a high-level authoring tool for domain experts.

With the framework developed, we achieved our goals concerning the development of a domain independent framework that supports the creation of personalized multimedia content independent of the final presentation format. Its design allows to "just" use the functionality it provides, for example, the access to media data, associated meta data, and user profile information, as well as the generation of the personalized multimedia content in standardized presentation formats. Hence, the framework relieves the developers of personalized multimedia applications from common tasks needed for content personalization, that is, personalized content selection, composition functionality, and presentation generation, and lets them concentrate on their application-specific job. However, the framework is also designed to be extensible with regard to application-specific personalization functionality, for example, by an application-specific personalized multimedia composition functionality. With the applications in the field of tourism and sports news we illustrated the usage of the framework in different domains and showed how the framework easily allows one to dynamically create personalized multimedia content for different user contexts and devices.

The framework has been designed not to become yet another framework but to base on and integrate previous and existing research in the field. The design has been based on our long-term experience in advanced multimedia composition models and an extensive study of previous and ongoing related approaches. Its interfaces and extensibility explicitly allow not only to extend the framework's functionality but to embrace existing solutions of other (research) approaches in the field. The dynamically created personalized multimedia content needs semantically rich annotated content in the respective media databases. In turn, the newly created content itself not only provides users with personalized multimedia information but

at the same time forms a new, semantically even richer multimedia content that can be retrieved and reused. The composition operators provide common multimedia composition functionality but also allow the integration of very specific operators. The explicit decision for presentation independence by a comprehensive internal composition model makes the framework both independent of any specific presentation format and prepares it for future formats to come.

Even though a software engineering approach towards dynamic creation of personalized multimedia content may not be the obvious one, we are convinced that our framework fills the gap between dynamic personalization support based on abstract data models, constraints or rules, and application-specific programming of personalized multimedia applications. With the MM4U framework, we are contributing to a new but obvious research challenge in the field of multimedia research, that is, the shift from tools for the manual creation of static multimedia content towards techniques for the dynamic creation of, respectively, context-aware and personalized multimedia content, which is needed in many application fields. Due to its domain independence, MM4U can be used by arbitrary personalized multimedia applications, each application applying a different configuration of the framework. Consequently, for providers of applications the framework approach supports a cheaper and quicker development process and by this contributes to a more efficient personalized multimedia content engineering.

REFERENCES

3rd Generation Partnership Project (2003a). *TS 26.234; Transparent end-to-end packet-switched streaming service; protocols and codecs (Release 5).* Retrieved December 19, 2003, from *http://www.3gpp.org/ftp/Specs/html-info/26234.htm*

3rd Generation Partnership Project (2003b). TS 26.246; *Transparent end-to-end packet-switched streaming service: 3GPP SMIL language profile (Release 6).* Retrieved December 19, 2003, from *http://www.3gpp.org/ftp/Specs/html-info/26246.htm*

Ackermann, P. (1996). *Developing object oriented multimedia software: Based on MET++ application framework*. Heidelberg, Germany: dpunkt.

Adobe Systems, Inc., USA (2001). *Adobe SVG Viewer*. Retrieved February 25, 2004, from *http://www.adobe.com/svg/*

Agnihotri, L., Dimitrova, N., Kender, J., & Zimmerman, J. (2003). Music videos miner. In *ACM Multimedia*.

Allen, J. F. (1983, November). Maintaining knowledge about temporal intervals. In *Commun. ACM*, *25*(11).

Amazon, Inc., USA (1996-2004). *Amazon.com*. Retrieved February 20, 2004, from *http://www.amazon.com/*

Andersson, O., Axelsson, H., Armstrong, P., Balcisoy, S., et al. (2004a). *Mobile SVG profiles: SVG Tiny and SVG Basic. W3C recommendation 25/03/2004*. Retrieved June 10, 2004, from *http://www.w3.org/TR/SVGMobile12/*

Andersson, O., Axelsson, H., Armstrong, P., Balcisoy, S., et al. (2004b). *Scalable vector graphics (SVG) 1.2 specification. W3C working draft 05/10/2004.* Retrieved June 10, 2004, from *http://www.w3c.org/Graphics/SVG/*

André, E. (1996). WIP/PPP: Knowledge-based methods for fully automated multimedia authoring. In *Proceedings of the EUROMEDIA'96*. London.

André, E., & Rist, T. (1995). Generating coherent presentations employing textual and visual material. In *Artif. Intell. Rev, 9*(2-3). Kluwer Academic Publishers.

André, E., & Rist, T. (1996, August). Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96),* Portland, Oregon.

Arndt, T. (1999, June). The evolving role of software engineering in the production of multimedia applications. In *IEEE International Conference on Multimedia Computing and Systems Volume 1,* Florence, Italy.

Ayars, J., Bulterman, D., Cohen, A., Day, K., Hodge, E., Hoschka, P., et al. (2001). *Synchronized multimedia integration language (SMIL 2.0) specification. W3C Recommendation 08/07/2001.* Retrieved February 23, 2004, from *http://www.w3c.org/AudioVideo/*

Beckett, D., & McBride, B. (2003). *RDF/XML syntax specification (revised). W3C recommendation 15/12/2003.* Retrieved February 23, 2004: *http://www.w3c.org/RDF/*

Bohrer, K., & Holland, B. (2004). *Customer profile exchange (cpexchange) specification – version 1.0, 20/10/2000.* Retrieved January 27, 2004, from *http://www.cpexchange.org/standard/cpexchangev1_0F.zip*

Boll, S. (2003, July). Vienna 4 U - What Web services can do for personalized multimedia applications. In *Proceedings of the Seventh Multi-Conference on Systemics Cybernetics and Informatics (SCI 2003),* Orlando, Florida, USA.

Boll, S., & Klas, W. (2001). ZYX - A multimedia document model for reuse and adaptation. In *IEEE Transactions on Knowledge and Data Engineering, 13*(3).

Boll, S., Klas, W., & Wandel, J. (1999, November). A cross-media adaptation strategy for multimedia presentations. *Proc. of the ACM Multimedia Conf. '99, Part 1,* Orlando, Florida, USA.

Boll, S., Klas, W., Heinlein, C., & Westermann, U. (2001, August). *Cardio-OP - Anatomy of a multimedia repository for cardiac surgery.* Technical Report TR-2001301, University of Vienna, Austria.

Boll, S., Klas, W., & Westermann, U. (2000, August). Multimedia Document Formats - Sealed Fate or Setting Out for New Shores? In *Multimedia - Tools and Applications, 11*(3).

Boll, S., Krösche, J., & Scherp, A. (2004, September). Personalized multimedia meets location-based services. In *Proceedings of the "Multimedia-Informationssysteme" Workshop associated with the 34th annual meeting of the German Society of Computing Science*, Ulm, Germany.

Boll, S., Krösche, J., & Wegener, C. (2003, August). Paper chase revisited - A real world game meet hypermedia (short paper). In *Proc. of the Intl. Conference on Hypertext (HT03),* Nottingham, UK.

Boll, S., & Westermann, U. (2003, November). MediÆther - An event space for context-aware multimedia experiences. In *Proc. of International ACM SIGMM Workshop on Experiential Telepresence*, Berkeley, CA., USA.

Bordegoni, M., Faconti, G., Feiner, S., Maybury, M. T., Rist, T., Ruggieri, S., et al. (1997, December). A standard reference model for intelligent multimedia presentation systems. In *ACM Computer Standards & Interfaces, 18*(6-7).

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction, 6*(2-3).

Bugaj, S., Bulterman, D., Butterfield, B., Chang, W., Fouquet, G., Gran, C., et al. (1998). *Synchronized multimedia integration language (SMIL 1.0) specification. W3C Recommendation 06/15/1998.* Retrieved June 10, 2004, from *http://www.w3.org/TR/REC-smil/*

Bulterman, D. C. A., van Rossum, G., & van Liere, R. (1991). A structure of transportable,

dynamic multimedia documents. In *Proceedings of the Summer 1991 USENIX Conf.,* Nashville, TN, USA.

Chen, G., & Kotz, D. (2000). *A survey of context-aware mobile computing research.* Technical Report TR2000-381. Dartmouth University, Department of Computer Science,

Click2learn, Inc, USA (2001-2002). *Toolbook – Standards-based content authoring.* Retrieved February 6, 2004 from *http://home.click2learn.com/en/toolbook/index.asp*

CWI (2004). *The cuypers multimedia transformation engine*. Amsterdam, The Netherlands. Retrieved February 25, 2004, on *http://media.cwi.nl:8080/demo/*

De Bra, P., Aerts, A., Berden, B., De Lange, B., Rousseau, B., Santic, T., Smits, D., & Stash, N. (2003, August). AHA! The Adaptive Hypermedia Architecture. *Proceedings of the ACM Hypertext Conference,* Nottingham, UK.

De Bra, P., Aerts, A., Houben, G.-J., & Wu, H. (2000). Making general-purpose adaptive hypermedia work. In *Proc. of the AACE WebNet Conference,* San Antonio, Texas.

De Bra, P., Aerts, A., Smits, D., & Stash, N. (2002a, October). AHA! version 2.0: More adaptation flexibility for authors. In *Proc. of the AACE ELearn2002 Conf.*

De Bra, P., Brusilovsky, P., & Conejo, R. (2002b, May). *Proc. of the Second Intl. Conf. for Adaptive Hypermedia and Adaptive Web-Based Systems,* Malaga, Spain, Springer LNCS 2347.

De Bra, P., Brusilovsky, P., & Houben, G.-J. (1999a, December). Adaptive hypermedia: From systems to framework. *ACM Computing Surveys, 31*(4).

De Bra, P., Houben, G.-J., & Wu, H. (1999b). AHAM: A dexter-based reference model for adaptive hypermedia. In *Proceedings of the 10th ACM Conf. on Hypertext and hypermedia: returning to our diverse roots,* Darmstadt, Germany.

De Carolis, B., de Rosis, F., Andreoli, C., Cavallo, V., De Cicco, M L (1998). The Dynamic Generation of Hypertext Presentations of Medical Guidelines. *The New Review of Hypermedia and Multimedia, 4.*

De Carolis, B., de Rosis, F., Berry, D., & Michas, I. (1999). Evaluating plan-based hypermedia generation. In *Proc. of European Workshop on Natural Language Generation,* Toulouse, France.

de Rosis, F., De Carolis, B., & Pizzutilo, S. (1999). Software documentation with animated agents. In *Proc. of the 5th ERCIM Workshop on User Interfaces For All,* Dagstuhl, Germanny.

Dublin Core Metadata Initiative (1995-2003). *Expressing simple Dublin Core in RDF/XML,1995-2003.* Retrieved February 2, 2004, from *http://dublincore.org/documents/2002/07/31/dcmes-xml/*

Duda, A., & Keramane, C. (1995). Structured temporal composition of multimedia data. *Proceedings of the IEEE International Workshop Multimedia-Database-Management Systems.*

Duke, D. J., Herman, I., & Marshall, M. S. (1999). *PREMO: A framework for multimedia middleware: Specification, rationale, and java binding.* New York: Springer.

Echiffre, M., Marchisio, C., Marchisio, P., Panicciari, P., & Del Rossi, S. (1998, January-March). MHEG-5 – Aims, concepts, and implementation issues. In *IEEE Multimedia.*

Egenhofer, M. J., & Franzosa, R. (1991, March). Point-Set Topological Spatial Relations. *Int. Journal of Geographic Information Systems, 5*(2).

Elhadad, M., Feiner, S., McKeown, K., & Seligmann, D. (1991). Generating customized text and graphics in the COMET explanation testbed. In *Proc. of the 23rd Conference on Winter Simulation. IEEE Computer Society,* Phoenix, Arizona, USA.

Engels, G., Sauer, S., & Neu, B. (2003, October). Integrating software engineering and user-centred

design for multimedia software developments. In *Proc. IEEE Symposia on Human-Centric Computing Languages and Environments - Symposium on Visual/Multimedia Software Engineering,* Auckland, New Zealand. IEEE Computer Society Press.

Exor International Inc. (2001-2004). eSVG: *Embedded SVG.* Retrieved February 12, 2004, from *http://www.embedding.net/eSVG/english/overview/overview_frame.html*

Fink, J., Kobsa, A., & Schreck, J. (1997). Personalized hypermedia information through adaptive and adaptable system features: User modeling, privacy and security issues. In A. Mullery, M. Besson, M. Campolargo, R. Gobbi, & R. Reed (Eds.), *Intelligence in services and networks: Technology for cooperative competition*. Berlin: Springer.

Foundation for Intelligent Physical Agents (2002). *FIPA device ontology specification, 2002.* Retrieved January 23, 2004, from *http://www.fipa.org/specs/fipa00091/*

Gaggi, O., & Celentano, A. (2002). A visual authoring environment for prototyping multimedia presentations. In *Proceedings of the IEEE Fourth International Symposium on Multimedia Software Engineering.*

Girgensohn, A., Bly, S., Shipman, F., Boreczky, J., & Wilcox, L. (2001). Home video editing made easy – Balancing automation and user control. In *Proc. of the Human-Computer Interaction,* Tokyo, Japan.

Girgensohn, A., Shipman, F., & Wilcox, L. (2003, November). Hyper-Hitchcock: Authoring Interactive Videos and Generating Interactive Summaries. In *Proc. ACM Multimedia*.

Greiner, C., & Rose, T. (1998, November). A Web based training system for cardiac surgery: The role of knowledge management for interlinking information items. In *Proc. The World Congress on the Internet in Medicine,* London.

Hardman, L. (1998, March). *Modeling and Authoring Hypermedia Documents.* Doctoral dissertation, University of Amsterdam, The Netherlands.

Hardman, L., Bulterman, D. C. A., & van Rossum, G. (1994b, February). The Amsterdam Hypermedia Model: Adding time and context to the Dexter Model. In *Comm. of the ACM, 37*(2).

Hardman, L., van Rossum, G., Jansen, J., & Mullender, S. (1994a). CMIFed: A transportable hypermedia authoring system. In *Proc. of the Second ACM International Conference on Multimedia,* San Francisco.

Hirzalla, N., Falchuk, B., & Karmouch, A. (1995). A temporal model for interactive multimedia scenarios. In *IEEE Multimedia, 2*(3).

Hunter, J. (1999, October). *Multimedia metadata schemas.* Retrieved June 16, 2004, from *http://www2.lib.unb.ca/ Imaging_docs/IC/schemas.html*

IBM Corporation, USA. (2004a). *IBM research – Video semantic summarization systems.* Retrieved June 15, 2004, from *http://www.research.ibm.com/MediaStar/VideoSystem.html#Summarization%20Techniques*

IBM Corporation, USA. (2004b). *QBIC home page*. Retrieved June 16, 2004, from *http://wwwqbic.almaden.ibm.com/*

INRIA (2003). *PocketSMIL 2.0*. Retrieved February 24, 2004, from *http://opera.inrialpes.fr/pocketsmil/*

International Organisation for Standardization (1996). *ISO 13522-5, information technology – Coding of multimedia and hypermedia information, Part 5: Support for base-level interactive applications.* Geneva, Switzerland: International Organisation for Standardization.

ISO/IEC (1999, July). *JTC 1/SC 29/WG 11, MPEG-7: Context, Objectives and Technical Roadmap,*

V.12. ISO/IEC Document N2861. Geneva, Switzerland: Int. Organisation for Standardization/Int. Electrotechnical Commission.

ISO/IEC (2001a, November). JTC 1/SC 29/WG 11. *Informationtechnology—Multimedia content description interface—Part 1: Systems.* ISO/IEC Final Draft International Standard 15938-1:2001. Geneva, Switzerland: Int. Organisation for Standardization/Int. Electrotechnical Commission.

ISO/IEC (2001b, September). JTC 1/SC 29/WG 11. *Information technology—Multimedia content description interface—Part 2: Description definition language.* ISO/IEC Final Draft Int. Standard 15938-2:2001. Geneva, Switzerland: Int. Organisation for Standardization/Int. Electrotechnical Commission.

ISO/IEC (2001c, July). JTC 1/SC 29/WG 11. *Information technology—Multimedia content description interface—Part 3: Visual.* ISO/IEC Final Draft Int. Standard 15938-3:2001. Geneva, Switzerland: Int. Organisation for Standardization/Int. Electrotechnical Commission.

ISO/IEC (2001d, June). JTC 1/SC 29/WG 11. *Information technology—Multimedia content description interface—Part 4: Audio.* ISO/IEC Final Draft Int. Standard 15938-4:2001. Geneva, Switzerland: Int. Organisation for Standardization/Int. Electrotechnical Commission.

ISO/IEC (2001e, October). JTC 1/SC 29/WG 11. *Information technology—Multimedia content description interface—Part 5: Multimedia description schemes.* ISO/IEC Final Draft Int. Standard 15938-5:2001. Geneva, Switzerland: Int. Organisation for Standardization/Int. Electrotechnical Commission.

Jourdan, M., Layaïda, N., Roisin, C., Sabry-Ismaïl, L., & Tardif, L. (1998). *Madeus, and authoring environment for interactive multimedia documents.* ACM Multimedia.

Kim, M., Wood, S., & Cheok, L.-T. (2000, November). Extensible MPEG-4 textual format (XMT). In *Proc. of the 8th ACM Multimedia Conf.*, Los Angeles.

Klas, W., Greiner, C., & Friedl, R. (1999, July). Cardio-OP: Gallery of cardiac surgery. IEEE *International Conference on Multimedia Computing and Systems (ICMS 99).* Florence, July.

Klyne, G., Reynolds, F., Woodrow, C., Ohto, H., Hjelm, J., Butler, M. H., & Tran, L. (2003). *Composite capability/preference profile (CC/PP): Structure and vocabularies* - W3C Working Draft 25/03/2003.

Kopf, S., Haenselmann, T., Farin, D., & Effelsberg, W. (2004). Automatic generation of summaries for the Web. In *Proceedings Electronic Imaging 2004.*

Lemlouma, T., & Layaïda, N. (2003, June). Media resources adaptation for limited devices. In *Proc. of the Sevent ICCC/IFIP International Conference on Electronic Publishing ELPUB 2003*, Universidade deo Minho, Portugal.

Lemlouma, T., & Layaïda, N. (2004, January). Context-aware adaptation for mobile devices. *IEEE International Conference on Mobile Data Management*, Berkeley, California, USA.

Little, T. D. C., & Ghafoor, A. (1993). Interval-based conceptual models for time-dependent multimedia data. In *IEEE Transactions on Knowledge and Data Engineering, 5*(4).

Macromedia, Inc., USA (2003, January). *Using Authorware 7.* [Computer manual]. Available from *http://www.macromedia.com/software/authorware/*

Macromedia, Inc., USA (2004). *Macromedia.* Retrieved June 15, 2004, from *http://www.macromedia.com/*

McKeown, K., Robin, J., & Tanenblatt, M. (1993). Tailoring lexical choice to the user's vocabulary in multimedia explanation generation. In *Proc. of the 31st conference on Association for Computational Linguistics*, Columbus, Ohio.

Oldenettel, F., & Malachinski, M. (2003, May). The LEBONED metadata architecture. In *Proc. of the 12th International World Wide Web Conference*, Budapest, Hungary (pp. S.207-216). ACM Press, Special Track on Education.

Open Mobile Alliance (2003). *User agent profile (UA Prof).* – 20/05/2003. Retrieved February 10, 2004, from *http://www.openmobilealliance.org/*

Oratrix (2004). GRiNS for SMIL Homepage. Retrieved February 23, 2004, from *http://www.oratrix.com/GRiNS*

Papadias, D., & Sellis, T. (1994, October). Qualitative representation of spatial knowledge in two-dimensional space. In *VLDB Journal, 3*(4).

Papadias, D., Theodoridis, Y., Sellis, T., & Egenhofer, M. J. (1995, March). Topological relations in the world of minimum bounding rectangles: A study with R-Trees. In *Proc. of the ACM SIGMOD Conf. on Management of Data*, San Jose, California.

Pree, W. (1995). *Design patterns for object-oriented software development.* Boston: Addison-Wesley.

Rabin, M. D., & Burns, M. J. (1996). Multimedia authoring tools. In *Conference Companion on Human Factors in Computing Systems*, Vancouver, British Columbia, Canada, ACM Press.

Raggett, D., Le Hors, A., & Jacobs, I. (1998). *HyperText markup language (HTML) version 4.0.* W3C Recommendation, revised on 04/24/1998. Retrieved February 20, 2004, from *http://www.w3c.org/MarkUp/*

RealNetworks (2003). *RealOne Player.* Retrieved February 25, 2004, from *http://www.real.com/*

Rossi, G., Schwabe, D., & Guimarães, R. (2001, May). Designing personalized Web applications. In *Proceedings of the tenth World Wide Web (WWW) Conference*, Hong Kong. ACM.

Rout, T. P., & Sherwood, C. (1999, May). Software engineering standards and the development of multimedia-based systems. In *Fourth IEEE International Symposium and Forum on Software Engineering Standards.* Curitiba, Brazil.

Scherp, A., & Boll, S. (2004a, March). *MobileMM4U - Framework support for dynamic personalized multimedia content on mobile systems.* In Multikonferenz Wirtschaftsinformatik 2004, special track on Technologies and Applications for Mobile Commerce.

Scherp, A., & Boll, S. (2004b, October). Generic support for personalized mobile multimedia tourist applications. Technical demonstration for the *ACM Multimedia Conference*, New York, USA.

Scherp, A., & Boll, S. (2005, January). Paving the last mile for multi-channel multimedia presentation generation. In *Proceedings of the 11th International Conference on Multimedia Modeling*, Melbourne, Australia.

Schmitz, P., Yu, J., & Santangeli, P. (1998). *Timed interactive multimedia extensions for HTML (HTML+TIME).* W3C, version 09/18/1998. Retrieved February 20, 2004, from *http://www.w3.org/TR/NOTE-HTMLplusTIME*

Stash, N., Cristea, A., & De Bra, P. (2004). Authoring of learning styles in adaptive hypermedia. In *WWW'04 Education Track*. New York: ACM.

Stash, N., & De Bra, P. (2003, June). Building Adaptive Presentations with AHA! 2.0. *Proceedings of the PEG Conference*, Sint Petersburg, Russia.

Sun Microsystems, Inc. (2004). *Java media framework API.* Retrieved February 15, 2004, from *http://java.sun.com/products/java-media/jmf/index.jsp*

Szyperski, C., Gruntz, D., & Murer, S. (2002). *Component software: Beyond object-oriented programming* (2nd ed.). Boston: Addison-Wesley.

van Ossenbruggen, J.R., Cornelissen, F.J., Geurts, J.P.T.M., Rutledge, L.W., & Hardman, H.L. (2000, December). *Cuypers: A semiautomatic hypermedia presentation system*. Technical Report INS-R0025. CWI, The Netherlands.

van Rossum, G., Jansen, J., Mullender, S., & Bulterman, D. C. A. (1993). CMIFed: A presentation environment for portable hypermedia documents. In *Proc. of the First ACM International Conference on Multimedia,* Anaheim, California.

Villard, L. (2001, November). Authoring transformations by direct manipulation for adaptable multimedia presentations. In *Proceeding of the ACM Symposium on Document Engineering,* Atlanta, Georgia.

Wahl, T., & Rothermel, K. (1994, May). Representing time in multimedia systems. In *Proc. IEEE Int. Conf. on Multimedia Computing and Systems,* Boston.

Wu, H., de Kort, E., & De Bra, P. (2001). Design issues for general-purpose adaptive hypermedia systems. In *Proc. of the 12th ACM Conf. on Hypertext and Hypermedia,* Århus, none, Denmark.

Yahoo!, Inc. (2002). *MyYahoo!*. Retrieved February 17, 2004, from *http://my.yahoo.com/*

This work was previously published in Managing Multimedia Semantics, edited by U. Srinivasan & S. Nepal, pp. 246-287, copyright 2006 by IRM Press (an imprint of IGI Global).

Chapter 2.3
EMMO:
Tradeable Units of Knowledge-Enriched Multimedia Content

Utz Westermann
University of Vienna, Austria

Sonja Zillner
University of Vienna, Austria

Karin Schellner
ARC Research Studio Digital Memory Engineering, Vienna, Austria

Wolfgang Klas
University of Vienna and ARC Research Studio Digital Memory Engineering, Vienna, Austria

ABSTRACT

Current semantic approaches to multimedia content modeling treat the content's media, the semantic description of the content, and the functionality performed on the content, such as rendering, as separate entities, usually kept on separate servers in separate files or databases and typically under the control of different authorities. This separation of content from its description and functionality hinders the exchange and sharing of content in collaborative multimedia application scenarios. In this chapter, we propose Enhanced Multimedia MetaObjects (Emmos) as a new content modeling formalism that combines multimedia content with its description and functionality. Emmos can be serialized and exchanged in their entirety — covering media, description, and functionality — and are versionable, thereby establishing a suitable foundation for collaborative multimedia applications. We outline a distributed infrastructure for Emmo management and illustrate the benefits and usefulness of Emmos and this infrastructure by means of two practical applications.

INTRODUCTION

Today's multimedia content formats such as HTML (Raggett et al., 1999), SMIL (Ayars et al., 2001), or SVG (Ferraiolo et al., 2003) primarily encode the presentation of content but not the information content conveys. But this *presentation-oriented* modeling only permits the hard-wired presentation of multimedia content exactly in the way specified; for advanced operations like retrieval and reuse, automatic composition, recommendation, and adaptation of content according to user interests, information needs, and technical infrastructure, valuable information about the semantics of content is lacking.

In parallel to research on the Semantic Web (Berners-Lee et al., 2001; Fensel, 2001), one can therefore observe a shift in paradigm towards a *semantic* modeling of multimedia content. The basic media of which multimedia content consists are supplemented with metadata describing these media and their semantic interrelationships. These media and descriptions are processed by stylesheets, search engines, or user agents providing advanced functionality on the content that can exceed mere hard-wired playback.

Current semantic multimedia modeling approaches, however, largely treat the content's basic media, the semantic description, and the functionality offered on the content as separate entities: the basic media of which multimedia content consists are typically stored on web or media servers; the semantic descriptions of these media are usually stored in databases or in dedicated files on web servers using formats like RDF (Lassila & Swick, 1999) or Topic Maps (ISO/IEC JTC 1/SC 34/WG 3, 2000); the functionality on the content is normally realized as servlets or stylesheets running in application servers or as dedicated software running at the clients such as user agents.

This inherent separation of media, semantic description, and functionality in semantic multimedia content modeling, however, hinders the realization of multimedia content sharing as well as collaborative applications which are gaining more and more importance, such as the sharing of MP3 music files (Gnutella, n.d.) or learning materials (Nejdl et al., 2002) or the collaborative authoring and annotation of multimedia patient records (Grimson et al., 2001). The problem is that exchanging content today in such applications simply means exchanging single media files. An analogous exchange of semantically modeled multimedia content would have to include content descriptions and associated functionality, which are only coupled loosely to the media and usually exist on different kinds of servers potentially under control of different authorities, and which are thus not easily moveable.

In this chapter, we give an illustrated introduction to Enhanced Multimedia MetaObjects (Emmo), a semantic multimedia content modeling approach developed with collaborative and content sharing applications in mind. Essentially, an Emmo constitutes a self-contained piece of multimedia content that merges three of the content's aspects into a single object: the *media aspect*, that is, the media which make up the multimedia content, the *semantic aspect* which describes the content, and the *functional aspect* by which an Emmo can offer meaningful operations on the content and its description that can be invoked and shared by applications. Emmos in their entirety — including media, content description, and functionality — can be *serialized* into bundles and are *versionable*: essential characteristics that enable their exchangeability in content sharing applications as well as the distributed construction and modification of Emmos in collaborative scenarios.

Furthermore, this chapter illustrates how we employed Emmos for two concrete collaborative and content sharing applications in the domains of cultural heritage and digital music archives.

The chapter is organized as follows: we begin with an overview of Emmos and show their difference to existing approaches for multimedia content

modeling. We then introduce the conceptual model behind Emmos and outline a distributed Emmo container infrastructure for the storage, exchange, and collaborative construction of Emmos. We then apply Emmos for the representation of multimedia content in two application scenarios. We conclude this paper with a summary and give an outlook to our current and future work.

BACKGROUND

In this section, we provide a basic understanding of the Emmo idea by means of an illustrating example. We show the uniqueness of this idea by relating Emmos to other approaches to multimedia content modeling in the field.

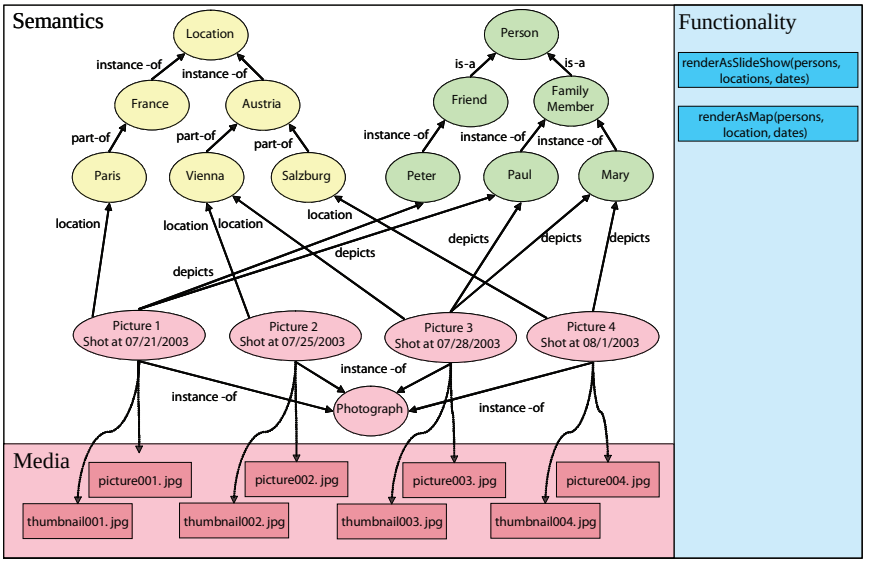
The Emmo Idea

The motivation for the development of the Emmo model is the desire to realize multimedia content sharing and collaborative applications on the basis of semantically modeled content but to avoid the limitations and difficulties of current semantic modeling approaches implied by their isolated treatment of media, semantic description, and content functionality.

Following an abstract vision originally formulated by (Reich et al., 2000), the essential idea behind Emmos is to keep semantic description and functionality tied to the pieces of content to which they belong, thereby creating self-contained units of semantically modeled multimedia content easier to exchange in content sharing and collaborative application scenarios. An Emmo coalesces the basic media of which a piece of multimedia content consists (i.e., the content's *media aspect*), the semantic description of these media (i.e., the content's *semantic aspect*), and functionality on the content (i.e., the content's *functional aspect*) into a single serializeable and versionable object.

Figure 1 depicts a sketch of a simplified Emmo, which models a small multimedia photo album of a holiday trip of an imaginary couple Paul and Mary and their friend Peter. The bottom of the figure illustrates how Emmos address the *media*

Figure 1. Aspects of an Emmo



aspect of multimedia content. An Emmo acts as a container of the basic media of which the content that is represented by the Emmo consists. In our example, the Emmo contains four JPEG images which constitute the different photographs of the album along with corresponding thumbnail images.

Media can be contained either by inclusion, that is, raw media data is directly embedded within an Emmo, or by reference via a URI if embedding raw media data is not feasible because of the size of media data or the media is a live stream.

An Emmo further carries a semantic description of the basic media it contains and the associations between them. This *semantic aspect*, illustrated to the upper left of Figure 1, makes an Emmo a unit of knowledge about the multimedia content it represents.

For content description, Emmos apply an expressive concept graph-like data model similar to RDF and Topic Maps. In this graph model, the description of the content represented by an Emmo is not performed directly on the media that are contained in the Emmo; instead, the model abstracts from physical media making it possible to subsume several media objects which constitute only different physical manifestations of logically one and the same medium under a single media node. This is a convenient way to capture alternative media. In Figure 1, for example, each media node Picture 1 … Picture 4 subsumes not only a photo but also its corresponding thumbnail image.

Apart from media, nodes can also represent abstract concepts. By associating an Emmo's media objects with such concepts, it is possible to create semantically rich descriptions of the multimedia content the Emmo represents. In Figure 1, for instance, it is expressed that the logical media nodes Picture 1 … Picture 4 constitute photos taken in Paris, Vienna, and Salzburg showing Peter and Paul, Paul and Mary, and Mary, respectively. The figure further indicates that nodes can be augmented with primitive attribute values for closer description: the pictures of the photo album are furnished with the dates at which they have been shot.

By associating concepts with each other, it is also possible to express domain knowledge within an Emmo. It is stated in our example that Peter, Paul, and Mary are Persons, that Paul and Mary are family members, that Peter is a friend, that Paris is located in France, and that Vienna and Salzburg are parts of Austria.

The Emmo model does not predefine the concepts, association types, and primitive attributes available for media description; these can be taken from arbitrary, domain-specific ontologies. While they thus constitute a very generic, flexible, and expressive approach to multimedia content model-

Figure 2. Emmo functionality

ing, Emmos are not ready-to-use formalism but require an agreed common ontology before they can be employed in an application.

Finally, Emmos also address the *functional aspect* of content. An Emmo can offer operations that can be invoked by applications in order to work with the content the Emmo represents in a meaningful manner. As shown to the top right of Figure 1, our example Emmo provides two operations supporting two different rendition options for the photo album, which are illustrated by the screenshots of Figure 2. As indicated by the left screenshot, the operation renderAsSlideshow()might know how to — given a set of persons, locations, as well as time periods of interest — render the photo album as a classic slideshow on the basis of the contained pictures and their semantic description by generating an appropriate SMIL presentation. As indicated by the right screenshot, the operation renderAsMap() might also know how to — given the same data — render the photo album as a map with thumbnails pointing to the locations where photographs have been taken by constructing an SVG graph.

One may think of many further uses of operations. For example, operations could also be offered for rights clearance, displaying terms of usage, and so forth.

Emmos have further properties: an Emmo can be *serialized* and shared in its entirety in a distributed content sharing scenario including its contained media, the semantic description of these media, and its operations. In our example, this means that Paul can accord Peter the photo album Emmo as a whole — for instance, via email or a file-sharing peer-to-peer infrastructure — and Peter can do anything with the Emmo that Paul can also do, including invoking its operations.

Emmos also support *versioning.* Every constituent of an Emmo is versionable, an essential prerequisite for applications requiring the distributed and collaborative authoring of multimedia content. This means that Peter, having received the Emmo from Paul, can add his own pictures to the photo album while Paul can still modify his local copy. Thereby, two concurrent versions of the Emmo are created. As the Emmo model is able to distinguish both versions, Paul can merge them into a final one when he receives Peter's changes.

Related Approaches

The fundamental idea underlying the concept of Emmos presented beforehand is that an Emmo constitutes an object unifying three different aspects of multimedia content, namely the media aspect, the semantic aspect, and the functional aspect. In the following, we fortify our claim that this idea is unique.

Interrelating basic media like single images and videos to form multimedia content is the task of multimedia document models. Recently, several standards for multimedia document models have emerged (Boll et al., 2000), such as HTML (Ragett et al., 1999), XHTML+SMIL (Newmann et al., 2002), HyTime (ISO/IEC JTC 1/SC 34/WG 3, 1997), MHEG-5 (ISO/IEC JTC 1/SC 29, 1997), MPEG-4 BIFS and XMT (Pereira & Ebrahimi, 2002), SMIL (Ayars et al., 2001), and SVG (Ferraiolo et al., 2003). Multimedia document models can be regarded as composite media formats that model the presentation of multimedia content by arranging basic media according to temporal, spatial, and interaction relationships. They thus mainly address the media aspect of multimedia content. Compared to Emmos, however, multimedia document models neither interrelate multimedia content according to semantic aspects nor do they allow providing functionality on the content. They rely on external applications like presentation engines for content processing.

As a result of research concerning the Semantic Web, a variety of standards have appeared that can be used to model multimedia content by describing the information it conveys on a semantic level, such as RDF (Lassila & Swick, 1999; Brickley & Guha, 2002), Topic Maps (ISO/IEC JTC 1/SC

34/WG 3, 2000), MPEG-7 (especially MPEG-7's graph tools for the description of content semantics (ISO/IEC JTC 1/SC 29/WG 11, 2001)), and Conceptual Graphs (ISO/JTC1/SC 32/WG 2, 2001). These standards clearly cover the semantic aspect of multimedia content. As they also offer means to address media within a description, they undoubtedly refer to the media aspect of multimedia content as well. Compared to Emmos, however, these approaches do not provide functionality on multimedia content. They rely on external software like database and knowledge base technology, search engines, user agents, and so forth, for the processing of content descriptions. Furthermore, media descriptions and the media described are separate entities — potentially scattered around different places on the Internet, created and maintained by different and unrelated authorities not necessarily aware of each other and not necessarily synchronized — whereas Emmos combine media and their semantic relationships into a single indivisible unit.

There exist several approaches that represent multimedia content by means of objects. Enterprise Media Beans (EMBs) (Baumeister, 2002) extend the Enterprise Java Beans (EJBs) architecture (Matena & Hapner, 1998) with predefined entity beans for the representation of basic media within enterprise applications. These come with rudimental access functionality but can be extended with arbitrary functionality using the inheritance mechanisms available to all EJBs. Though addressing the media and functional aspects of content, EMBs in comparison to Emmo are mainly concerned with single media content and not with multimedia content. Furthermore, EMBs do not offer any dedicated support for the semantic aspect of content.

Adlets (Chang & Znati, 2001) are objects that represent individual (not necessarily multimedia) documents. Adlets support a fixed set of predefined functionality which enables them to advertise themselves to other Adlets. They are thus content representations that address the media as well as the functional aspect. Different from Emmos, however, the functionality supported by Adlets is limited to advertisement and there is no explicit modeling of the semantic aspect.

Tele-Action Objects (TAOs) (Chang et al., 1995) are object representations of multimedia content that encapsulate the basic media of which the content consists and interlink them with associations. Though TAOs thus address the media aspect of multimedia content in a way similar to Emmos, they do not adequately cover the semantic aspect of multimedia content: only a fixed set of five association types is supported mainly concerned with temporal and spatial relationships for presentation purposes. TAOs can further be augmented with functionality. Such functionality is, in contrast to the functionality of Emmos, automatically invoked as the result of system events and not explicitly invoked by applications.

Distributed Active Relationships (Daniel et al., 1998) define an object model based on the Warwick Framework (Lagoze et al., 1996). In the model, Digital Objects (DOs), which are interlinked with each other by semantic relationships, act as containers of metadata describing multimedia content. DOs thus do not address the media aspect of multimedia content but focus on the semantic aspect. The links between containers can be supplemented with arbitrary functionality. As a consequence, DOs take account of the functional aspect as well. Different from Emmos, however, the functionality is not explicitly invoked by applications but implicitly whenever an application traverses a link between two DOs.

ENHANCED MULTIMEDIA META OBJECTS

Having motivated and illustrated the basic idea behind them, this section semiformally introduces the conceptual model underlying Emmos using UML class diagrams. A formal definition of this

Figure 3. Management of basic media in an Emmo

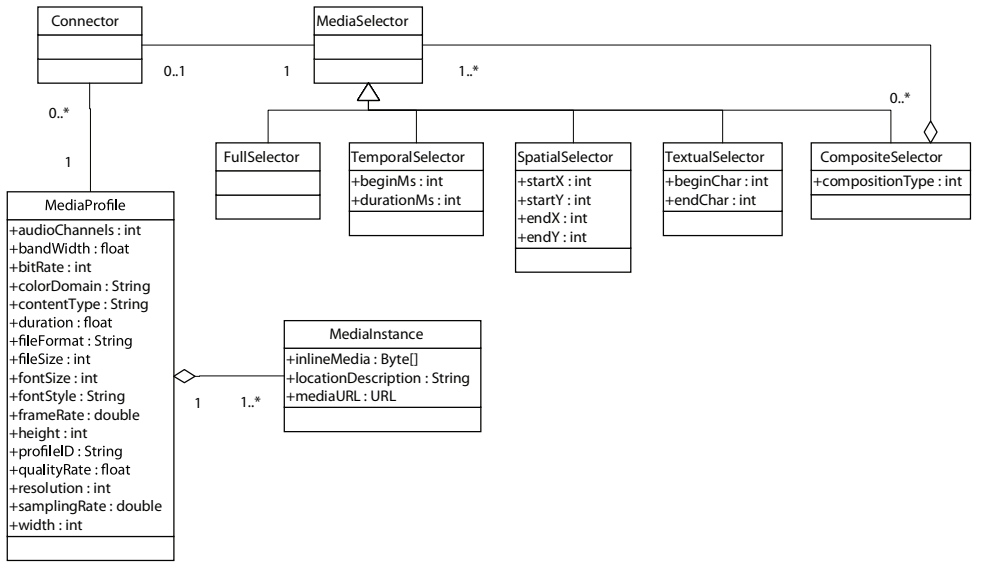
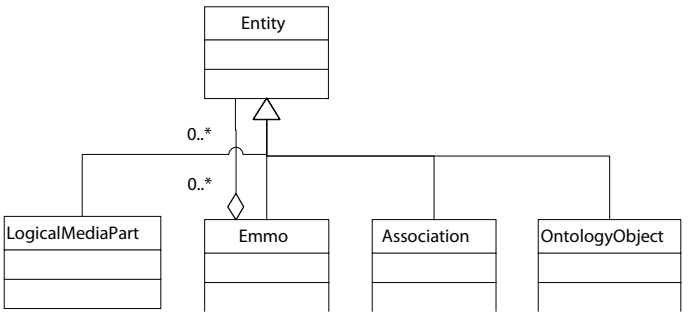


Figure 4. Semantic aspect of Emmos



model can be found in (Schellner et al., 2003). The discussion is oriented along the three aspects of multimedia content encompassed by Emmos: the media aspect, the semantic aspect, and the functional aspect.

Media Aspect

Addressing the media aspect of multimedia content, an Emmo encapsulates the basic media of which the content it represents is composed. Figure 3 presents the excerpt of the conceptual model which is responsible for this.

Closely following the MPEG-7 standard and its multimedia description tools (ISO/IEC JTC 1/SC 29/WG 11, 2001), basic media are modeled by *media profiles* (represented by the class MediaProfile in Figure 3) along with associated media instances (represented by the class MediaInstance). Media profiles hold low-level metadata describing physical characteristics of the media such as the storage format, file size, and so forth.; the media data itself is represented by *media instances*, each of which may directly embed the data in form of a byte array or, if that is not possible or feasible, address its storage lo-

cation by means of a URI. Moreover, if a digital representation is not available, a textual location description can be specified, for example the location of analog tapes in some tape archive. Figure 3 further shows that a media profile can have more than one media instances. In this way, an Emmo can be provided with information about alternative storage locations of media.

Basic media represented by media profiles and media instances are attached to an Emmo by means of a *connector* (see class Connector in Figure 3). A connector does not just address a basic medium via a media profile; it may also refer to a *media selector* (see base class MediaSelector) to address only a part of the medium. As indicated by the various subclasses of MediaSelector, it is possible to select media parts according to simple textual, spatial, temporal and textual criteria, as well as an arbitrary combination of these criteria (see class CompositeSelector). It is thus possible to address the upper right part of a scene in a digital video starting from second 10 and lasting until second 30 within an Emmo without having to extract that scene and put it into a separate media file using a video editing tool.

Semantic Aspect

Out of the basic media which it contains, an Emmo forges a piece of semantically modeled multimedia content by describing these media and their semantic interrelationships. The class diagram of Figure 4 gives an overview over the part of the Emmo model that provides these semantic descriptions. As one can see, the basic building blocks of the semantic descriptions, the so-called *entities*, are subsumed under the common base class Entity. The Emmo model distinguishes four kinds of entities: namely, *logical media parts*, *associations*, *ontology objects*, and *Emmos* themselves, represented by assigned subclasses. These four kinds of entities have a common nature but each extends the abstract notion of an entity with additional characteristic features.

Figure 5 depicts the characteristics that are common to all kinds of entities. Each entity is globally and uniquely identified by its OID, realized by means of a universal unique identifier (UUID) (Leach, 1998) which can be easily created even in distributed scenarios. To enhance human readability and usability, each entity is further augmented with additional attributes like a name and a textual description. Moreover, each entity holds information about its creator and its creation and modification date.

Figure 5 further expresses that entities may receive an arbitrary number of *types*. A type is a concept taken from an ontology and represented by an ontology object in the model. Types thus constitute entities themselves. By attaching types, an entity gets meaning and is classified in an application-dependent ontology. As mentioned before, the Emmo model does not come with a predefined set of ontology objects but instead relies on applications to agree on common ontology before the Emmo model can be used.

In the example of Figure 6, the entity Picture 3 of kind logical media part (depicted as a rectangle), which represents the third picture of our example photo album of the holiday trip introduced in the previous section, is an instantiation of the concepts "photograph" and "digital image", represented by the ontology objects photograph and digital image (each pictured by an octagon), respectively. The type relationships are indicated by dashed arrows.

For further description, the Emmo model also allows attaching arbitrary *attribute values* to entities (expressed by the class of the same name in the class diagram of Figure 5). Attribute values are simple attribute-value pairs, with the attributes being a concept of an application-dependent ontology represented by an ontology object entity, and the value being an arbitrary object suiting the type of the value. The rationale behind representing attributes by concepts of an ontology and not just by mere string identifiers is that this allows expressing constraints on the usage of attributes

Figure 5. Entity details

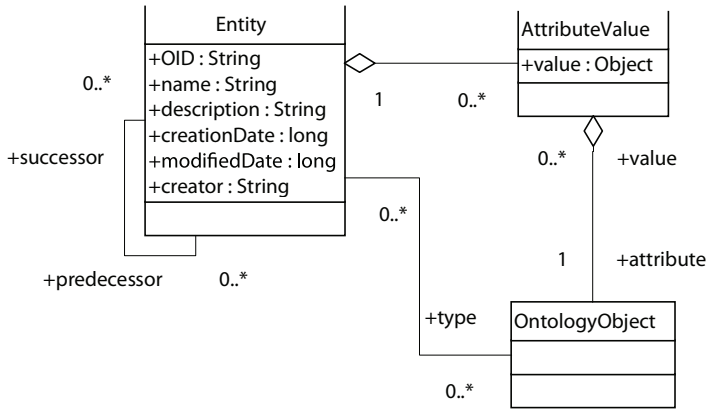
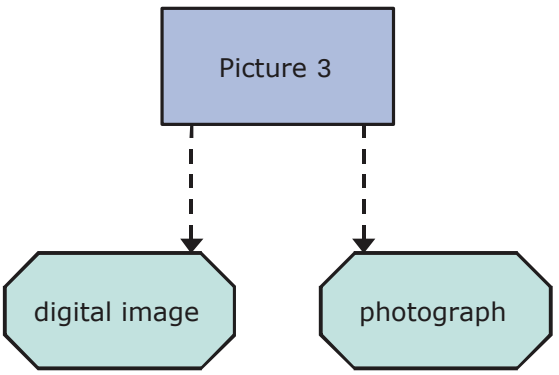


Figure 6. Entity with its types



within the ontology; for example, to which entity types attributes are applicable.

Figure 7 gives an example of attribute values. In the figure, it is stated that the third picture of the photo album was taken July 28, 2003, by attaching an attribute value "date=07/28/2003" to the entity Picture 3 representing that picture. The attribute "date" is modeled by the ontology object date and the value "07/28/2003" is captured by an object of a suitable date class (represented using the UML object notation).

As an essential prerequisite for the realization of distributed, collaborative multimedia applications in which multimedia content is simultaneously authored and annotated by different persons at different locations, the Emmo model provides intrinsic support for versioning. The class diagram of Figure 5 states that every entity is versionable and can have an arbitrary number of predecessor and successor versions, all of which have to be entities of the same kind as the original entity. Treating an entity's versions as entities on their own has several benefits: on the one hand, entities constituting versions of other entities have their own globally unique OID. Hence, different versions concurrently derived from one and the same entity at different sites can easily be distinguished without synchronization effort. On the other hand, different versions of an entity can be interrelated just like any other entities allowing one to establish comparative relationships between entity versions.

Figure 7. Entiy with an attribute value

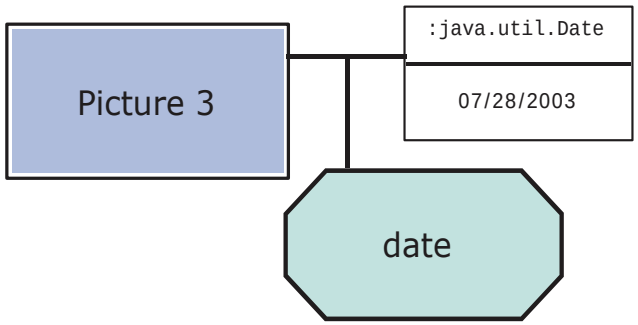


Figure 8. Versioning of an entity

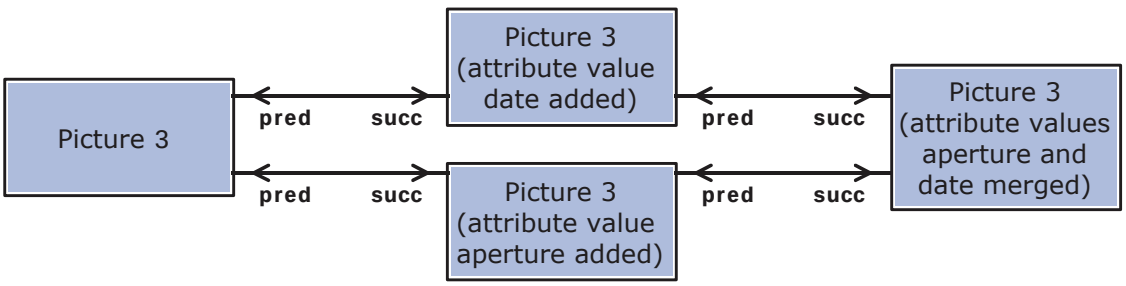


Figure 8 exemplifies a possible versioning of our example entity Picture 3. The original version of this logical part is depicted to the left of the figure. As expressed by the special arrows indicating the predecessor (pred) and the successor (succ) relationship between different versions of the same entity, two different successor versions of this original version were created, possibly by two different people at two different locations. One version augments the logical media part with a date attribute value to denote the creation date of the picture whereas the other provides an attribute value describing the aperture with which the picture was taken. Finally, as shown by the logical media part at the right side of the figure, these two versions were merged again into a fourth that now holds both attribute values.

Having explained the common characteristics shared by all entities, we are now able to introduce the peculiarities of the four concrete kinds of entities: logical media parts, ontology objects, associations, and Emmos.

Logical Media Parts

Logical media parts are entities that form the bridge between the semantic aspect and the media aspect of an Emmo. A logical media part represents a basic medium of which multimedia content consists on a logical level for semantic description, thereby providing an abstraction from the physical manifestation of the medium. According to the class diagram of Figure 9, logical media parts can refer to an arbitrary number of connectors — which we already know from our description of the media aspect of Emmo — permitting one to logically subsume alternative media profiles and instances representing different

media files in possibly different formats in possibly different storage locations under a common logical media part. The ID of the default profile to use is identified via the attribute masterProfileID. Since logical media parts do not need to have connectors associated with them, it is also possible to refer to media within Emmos which do not have a physical manifestation.

Ontology Objects

Ontology objects are entities that represent concepts of an ontology. We have already described how ontology objects are used to define entity types and to augment entities with attribute values. By relating entities such as logical media parts to ontology objects, they can be given a meaning. As it can be seen from the class diagram of Figure 10, the Emmo model distinguishes two kinds of ontology objects represented by two subclasses of OntologyObject: Concept and ConceptRef. Whereas an instance of Concept serves to represent a concept of an ontology that is fully captured within the Emmo model, ConceptRef allows one to reference concepts of ontologies specified in external ontology languages such as RDF Schema (Brickley & Guha, 2002). The latter is a pragmatic tribute to the fact that we have not developed an ontology language for Emmos yet and therefore rely on external languages for this purpose. References to concepts of external ontologies additionally need a special ID (ob-

Figure 9. Logical media parts

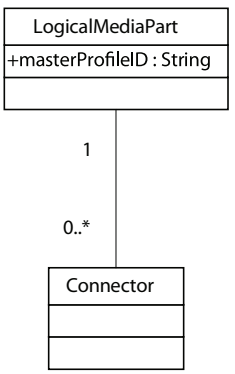


Figure 10. Ontology objects

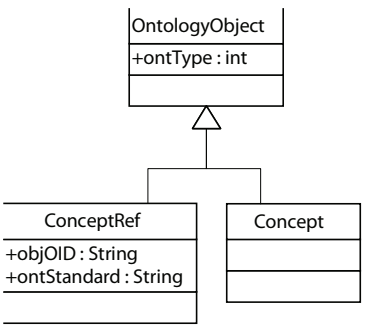
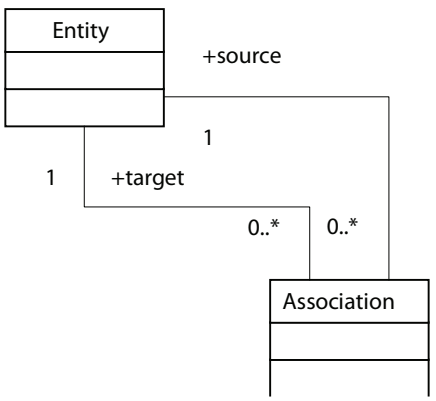


Figure 11. Association



jOID) uniquely identifying the external concept referenced and a label indicating the format of the ontology (ontStandard); for example, "RDF Schema".

Associations

Associations are entities that establish binary directed relationships between entities, allowing the creation of complex and detailed descriptions of the multimedia content represented by the Emmo. As one can see from Figure 11, each association has exactly one *source entity* and one *target entity*. The kind of semantic relationship represented by an association is defined by the association's *type* which is — like the types of other entities — an ontology object representing the concept that captures the type in an ontology. Different from other entities, however, an association is only permitted to have one type as it can express only a single kind of relationship.

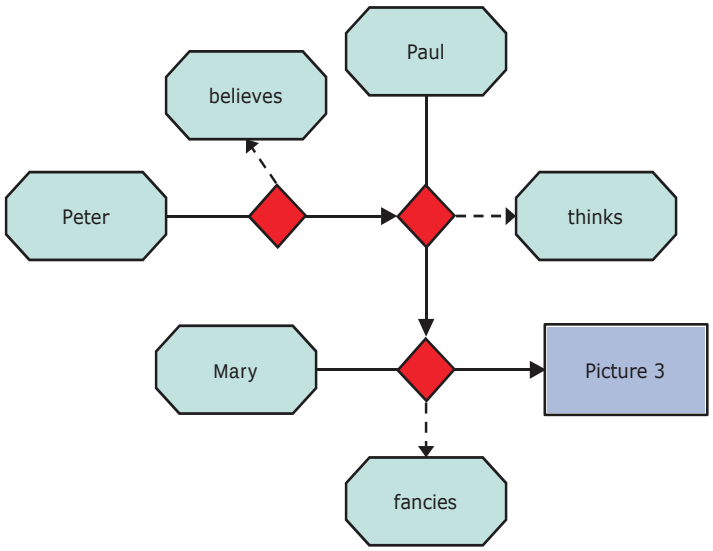
Since associations are first-class entities, they can take part as sources or targets in other associations like any other entities. This feature permits the creation of very complex content descriptions, as it facilitates the reification of statements ("statements about statements") within the Emmo model.

Figure 12 demonstrates how reification can be expressed. In the figure, associations are symbolized by a diamond shape, with solid arrows indicating the source and target of an association and dashed arrows indicating the association type. The example shown in this figure wants to express that "Peter believes that Paul thinks that Mary fancies Picture 3". The statement "Mary fancies Picture 3" is represented at the bottom of the figure by an association of type fancies that connects the ontology object Mary with the logical media part Picture3. Moreover, this association acts as target for another association having the type thinks and the source entity Paul, thereby making a statement about the statement "Mary fancies Picture 3". This reification is then further enhanced by attaching another statement to obtain the desired message.

Emmos

Emmos themselves, finally, constitute the fourth kind of entities. An Emmo is basically a container

Figure 12. Reification



that encapsulates arbitrary entities to form a semantically modeled piece of multimedia content (see the aggregation between the classes Emmo and Entity in the introductory outline of the model in Figure 4). As one and the same entity can be contained in more than one Emmo, it is possible to encapsulate different, context-dependent, and even contradicting, views onto the same content within different Emmo; as Emmo are first-class entities, they can be contained within other Emmos and take part in associations therein, allowing one to build arbitrarily nested Emmo structures for the logical organization of multimedia content. These are important characteristics especially useful for the authoring process, as they facilitate reuse of existing Emmos and the content they represent.

Figure 13 shows an example where a particular Emmo encapsulates another. In the figure, Emmos are graphically shown as ellipses. The example depicts an Emmo modeling a private photo gallery that up to the moment holds only a single photo album (again modeled by an Emmo): namely, the photo album of the journey to Europe we used as a motivating example in the section illustrating the Emmo idea. Via an association, this album is classified as "vacation" within the photo gallery. In the course of time, the photo gallery might become filled with additional Emmos representing further photo albums; for example, one that keeps the photos of a summer vacation in Spain. These Emmos can be related to each other. For example, an association might express that the journey to Europe took place before the summer vacation in Spain.

Functional Aspect

Emmos also address the functional aspect of multimedia content. Emmos may offer *operations* that realize arbitrary content-specific functionality which makes use of the media and descriptions provided with the media and semantic aspects of an Emmo and which can be invoked by applications working with content. The class diagram of Figure 14 shows how this is realized in the model. As expressed in the diagram, an Emmo may aggregate an arbitrary number of operations

Figure 13. Nested Emmo

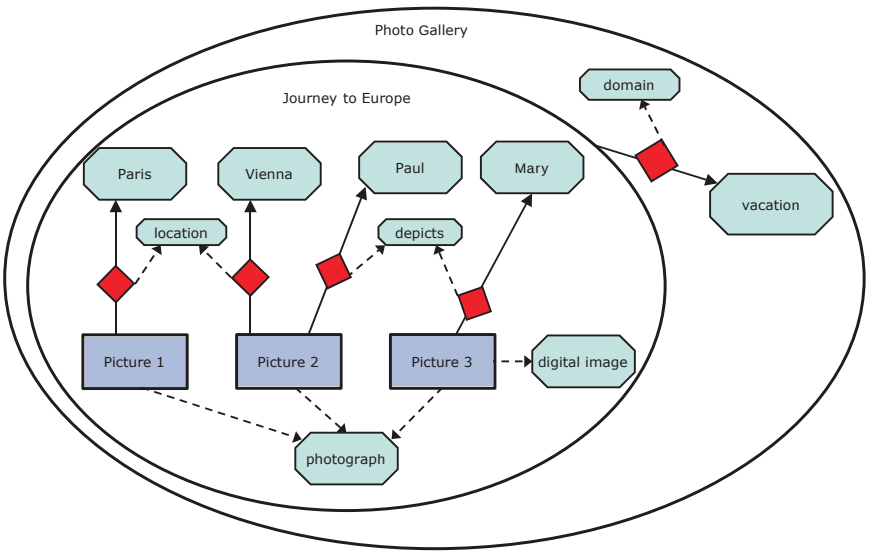


Figure 14. Emmo's functionality

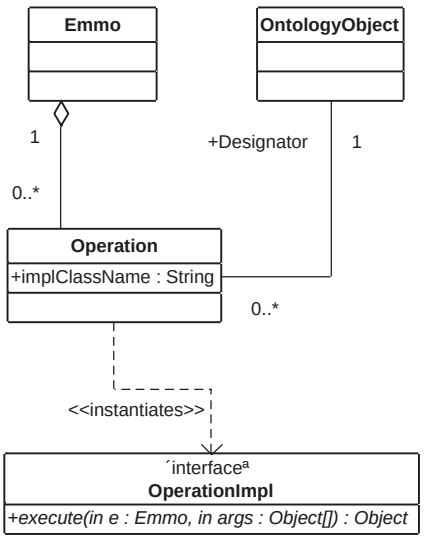
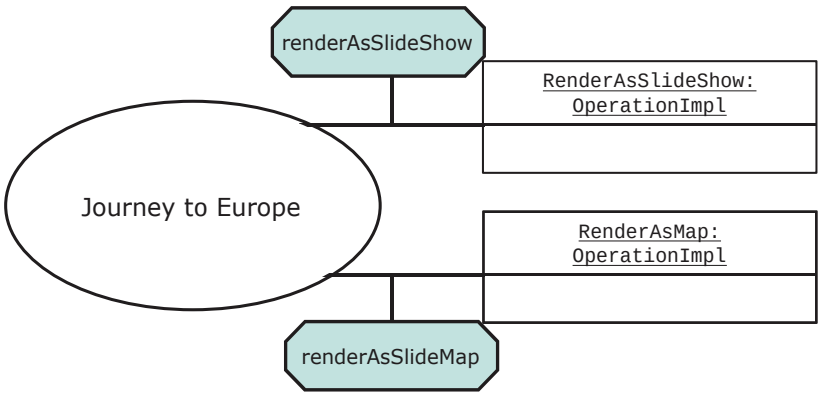


Figure 15. Example of Emmo operations



represented by the class of the same name. Each operation has a *designator*, that is, a name that describes its functionality, which is represented by an ontology object. Similar to attributes, the motivation behind using concepts of an ontology as operation designators instead of simple string identifiers is that this allows one to express restrictions on the usage of operations within an ontology; for example, the types of Emmo for which an operation is available, the types of the expected input parameters, and so forth.

The functionality of an operation is provided by a dedicated *implementation class* whose name is captured by an operation's implClassName attribute to permit the dynamic instantiation of the implementation class at runtime. There are not many restrictions for such an implementation class: the Emmo model merely demands that an

implementation class realizes the OperationImpl interface. OperationImpl enforces the implementation of a single method only: namely, the method execute() which expects the Emmo on which an operation is executed as its first parameter followed by a vector of arbitrary operation-dependent parameter objects. Execute() performs the desired functionality and, as a result, may return an arbitrary object.

Figure 15 once more depicts the Emmo modeling the photo album of the journey to Europe that we already know from Figure 13, but this time enriched with the two operations already envisioned in the second section: one that traverses the semantic description of the album returns an SMIL presentation that renders the album as a slide show, and another that returns an SVG presentation that renders the same album as a map. For both operations, two implementation classes are provided that are attached to the Emmo and differentiated via their designators renderAsSlideShow and renderAsMap.

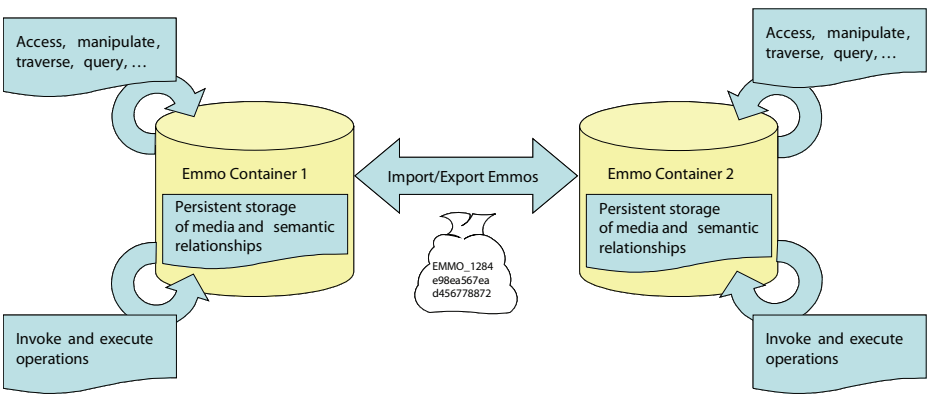
THE EMMO CONTAINER INFRASTRUCTURE

As an elementary foundation for the sharing and collaborative authoring of pieces of semantically modeled multimedia content on the basis of the Emmo model, we have implemented a distributed Emmo container infrastructure. Figure 16 provides an overview of this infrastructure, which we are going to describe in more detail in the following section.

Basically, an Emmo container provides a space where Emmos "live." Its main purpose is the management and persistent storage of Emmos. An Emmo container provides application programming interfaces that permit applications to fine-grainedly access, manipulate, traverse, and query the Emmos it stores. This includes the media aspect of an Emmo with its media profiles and instances, the semantic aspect with all its descriptional entities such as logical media parts, ontology objects, other Emmos, and associations, as well as the versioning relationships between those entities. Moreover, an Emmo container offers an interface to invoke and execute an Emmo's operations giving access to the functional aspect of an Emmo.

Emmo containers are not intended as a centralized infrastructure with a single Emmo container running at a server (although this is possible). Instead, it is intended to establish a decentralized infrastructure with Emmo containers of different scales and sizes running at each site that works with Emmos. Such a decentralized

Figure 16. Emmo container infrastructure



Emmo management naturally reflects the nature of content sharing and collaborative multimedia applications.

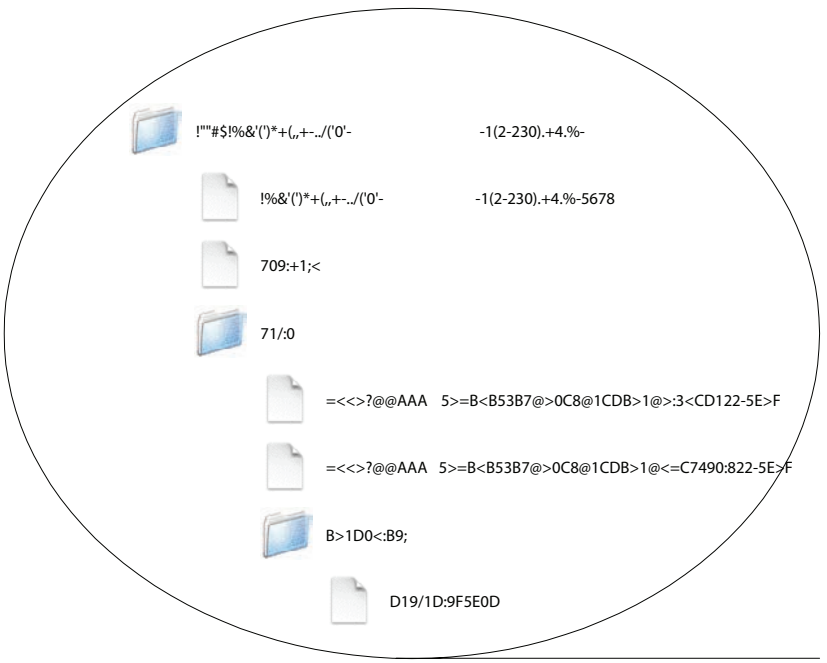
The decentralized approach has two implications. The first implication is that *platform independence* and *scalability* are important in order to support Emmo containers at potentially very heterogeneous sites ranging from home users to large multimedia content publishers with different operating systems, capabilities, and requirements.

For these reasons, we have implemented the Emmo containers in Java, employing the object-oriented DBMS ObjectStore for persistent storage. By Java, we obtain platform independence; by ObjectStore, we obtain scalability as there does not just exist a full-fledged database server implementation suitable for larger content providers, but also a code-compatible file-based in-process variant named PSEPro better suiting the limited needs of home users. It would have been possible to use a similarly scalable relational DBMS for persistent storage as well; we opted for an object-oriented DBMS, however, because of these systems' suitability for handling complex graph structures like Emmos.

The second implication of a decentralized infrastructure is that Emmos must be *transferable* between the different Emmo containers operated by users that want to share or collaboratively work on content. This requires Emmo containers to be able to completely export Emmos into bundles encompassing their media, semantic, and functional aspects, and to import Emmos from such bundles, which is explained in more detail in the following two subsections.

In the current state of implementation, Emmo containers are rather isolated components, requiring applications to explicitly initiate the import and export of Emmos and to manually transport Emmo bundles between different Emmo containers themselves. We are building a peer-to-peer infrastructure around Emmo containers that permits the transparent search for and transfer of Emmos across different containers.

Figure 17. Structure of an Emmo bundle



Exporting Emmos

An Emmo container can export an Emmo into a bundle whose overall structure is illustrated by Figure 17.

The bundle is basically a ZIP archive which captures all three aspects of an Emmo: the media aspect is captured by the bundle's media folder. The basic media files of which the multimedia content modeled by the Emmo consists are stored in this folder.

The semantic aspect is captured by a central XML file whose name is given the OID of the bundled Emmo. This XML file captures the semantic structure of the Emmo, thus describing all of the Emmo's entities, the associations between them, the versioning relationships, and so forth.

Figure 18 shows a fragment of such an XML file. It is divided into a <components> section declaring all entities and media profiles relevant for the current Emmo and a <links> section capturing all kinds of relationships between these entities and media profiles, such as types, associations, and so forth.

The functional aspect of an Emmo is captured by the bundle's operations folder in which the binary code of the Emmo's operations is stored. Here, our choice for Java as the implementation language for Emmo containers comes in handy again, as it allows us to transfer operations in form of JAR files with platform-independent bytecode even between heterogeneous platforms.

The export functionality can react to different application needs by offering several export variants: an Emmo can be exported with or without media included in the bundle, one can choose whether to also include media that are only referenced by URIs, the predecessor and successor versions of the contained entities can either be added to the bundle or omitted, and it can be decided whether to recursively export Emmos contained within an exported Emmo. The particular export variants chosen are recorded in the bundle's manifest file.

In order to implement these different export variants, an Emmo container distinguishes three different *modes* of how entities can be placed in a bundle:

- The *strong* mode is the normal mode for an entity. The bundle holds all information about an entity including its types, attribute values, immediate predecessor and successor versions, media profiles (in case of a logical media part), contained entities (in case of an Emmo), and so forth.
- The *hollow* mode is applicable to Emmos only. The hollow mode indicates that the bundle holds all information about an Emmo except the entities it contains. The hollow mode appears in bundles where it was chosen not to recursively export encapsulated Emmo. In this case, encapsulated Emmos receive the hollow mode; the entities encapsulated by those Emmos are excluded from the export.
- The *weak* mode indicates that the bundle contains only basic information about an entity, such as its OID, name, and description but no types, attribute values, and so forth. Weak mode entities appear in bundles that have been exported without versioning information. In this case, the immediate predecessor and successor versions of exported entities are placed into the bundle in weak mode; indirect predecessor and successor versions are excluded from the export.

The particular mode of an entity within a bundle is marked with the mode attribute in the entity's declaration in the bundle's XML file (see again Figure 18).

Importing Emmos

When importing an Emmo bundle exported in the way described in the previous subsection, an Emmo container essentially inserts all media files, entities, and operations included in the bundle into

Figure 18. Emmo XML representation

```
<?xml version="1.0" encoding="UTF-16"?>
<!-- Document created by org.cultos.storage.mdwb.exporter.MdwbXMLExporter -->
<emmo xmlns="http://www.cultos.org/emmos" xmlns:mpeg7="http://www.mpeg7.org/2001/MPEG-7_Schema"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.cultos.org/emmos http://www.cultos.org/emmos/XML/emmo.xsd">
      <components>
            <entities>
                  <entity xsi:type="LogicalMediaPart" mode="strong">
                        <oid>E1a8d252-f8f2e098bb-3c7cb04afdbd1144ba4d1ea866d93db2</oid>
                        <name>Beethoven's 5th Symphony</name>
                        <creationDate>19 November 2003 08:24:54  CET</creationDate>
                        <modifiedDate>19 November 2003 08:24:55  CET</modifiedDate>
                  </entity>
                  <entity xsi:type="Concept" mode="weak">
                        <oid>E1a8d252-f8f2d7a8a5-3c7cb04afdbd1144ba4d1ea866d93db2</oid>
                        <name>Classical music</name>
                        <creationDate>19 November 2003 08:15:09 CET</creationDate>
                        <modifiedDate>19 November 2003 08:15:09 CET</modifiedDate>
                        <ontologyType>0</ontologyType>
                  </entity>
                  .....
            </entities>
            <mediaProfiles/>
      </components>
      <links>
            <types>
                  <typeLink entity="E1a8d252-f8f2e095ed-3c7cb04afdbd1144ba4d1ea866d93db2"
                        type="E1a8d252-f8f2e0980f-3c7cb04afdbd1144ba4d1ea866d93db2"/>
                  <typeLink entity="E1a8d252-f8f2e095fc-3c7cb04afdbd1144ba4d1ea866d93db2"
                        type="E1a8d252-f8f2e098ef-3c7cb04afdbd1144ba4d1ea866d93db2"/>
                  <typeLink entity="E1a8d252-f8f2e098bb-3c7cb04afdbd1144ba4d1ea866d93db2"
                        type="E1a8d252-f8f2d7a8a5-3c7cb04afdbd1144ba4d1ea866d93db2"/>
            </types>
            <attributeValues/>
            <associations>
                  <assoLink association="E1a8d252-f8f2e0981a-3c7cb04afdbd1144ba4d1ea866d93db2"
                        sourceEntity="E1a8d252-f8f2e095fc3c7cb04afdbd1144ba4d1ea866d93db2"
                        targetEntity="E1a8d252-f8f2e098bb-3c7cb04afdbd1144ba4d1ea866d93db2"/>
            </associations>
            <connectors/>
            <predVersions/>
            <succVersions/>
            <encapsulations/>
      </links>
</emmo>
```

its local database. In order to avoid duplicates, the container checks whether an entity with the same OID or whether a media file or JAR file already exists in the local database before insertion. If a file already exists, the basic strategy of the importing container is that the local copy prevails.

However, the different export variants for Emmos and the different modes in which entities might occur in a bundle — as well as the fact that in a collaborative scenario Emmos might have been concurrently modified without creating new versions of entities — demand a more sophisticated handling of duplicate entities on the basis of a timestamp protocol. Depending on the modes of two entities with the same OID in the bundle, and the local database and the timestamps of both entities, essentially the following treatment is applied:

- A greater mode (weak < hollow < strong) in combination with a more recent timestamp

always wins. Thus, if the local entity has a greater mode and a newer timestamp, it prevails, and the entity in the bundle is ignored. Similarly, if the local entity has a lesser mode and an older timestamp, the entity in the bundle completely replaces the local entity in the database.

- If the local entity has a more recent timestamp but a lesser mode, additional data available for the entity in the bundle (entity types, attribute values, predecessor or successor versions, encapsulated entities in case of Emmos, or media profiles in case of logical media parts) complements the data of the local entity, thereby raising its mode.
- In case of same modes but a more recent timestamp of the entity in the bundle, the entity in the bundle completely replaces the local entity in the database.
- In case of same modes but a more recent timestamp of the entity in the local database, the entity in the database prevails and the entity in the bundle is ignored.

APPLICATIONS

Having introduced and described the Emmo approach to semantic multimedia content modeling and the Emmo container infrastructure, this section illustrates how these concepts have been practically applied in two concrete multimedia content sharing and collaborative applications. The first application named CULTOS is in the domain of cultural heritage and the second application introduces a semantic jukebox.

CULTOS

CULTOS is an European Union (EU)-funded project carried out from 2001 to 2003 with 11 partners from EU-countries and Israel1. It has been the task of CULTOS to develop a multimedia collaboration platform for authoring, managing, retrieving, and exchanging *Intertextual Threads* (ITTs) (Benari et al., 2002; Schellner et al., 2003) — knowledge structures that semantically interrelate and compare cultural artifacts such as literature, movies, artworks, and so forth. This platform enables the community of intertextual studies to create and exchange multimedia-enriched pieces of cultural knowledge that incorporate the community's different cultural backgrounds — an important contribution to the preservation of European cultural heritage.

ITTs are basically graph structures that describe semantic relationships between cultural artifacts. They can take a variety of forms, ranging from spiders over centipedes to associative maps, like the one shown in Figure 19.

The example ITT depicted in the figure highlights several relationships of the poem "The Fall" by Tuvia Ribner to other works of art. It states that the poem makes reference to the 3rd book of Ovid's "Metamorphoses" and that the poem is an ekphrasis of the painting "Icarus' Fall" by the famous Dutch painter Breugel.

The graphical representation of an ITT bears strong resemblance to well-known techniques for knowledge representation such as concept graphs or semantic nets, although it lacks their formal rigidity. ITTs nevertheless get very complex, as they commonly make use of constructs such as *encapsulation* and *reification* of statements that are challenging from the perspective of knowledge representation.

Encapsulation is intrinsic to ITTs because intertextual studies are not exact sciences. Certainly, the cultural and personal context of a researcher affects the kind of relationships between pieces of literature he discovers and are of value to him. As such different views onto a single subject are highly interesting to intertextual studies, ITTs themselves can be relevant subjects of discourse and thus be contained as first-class artifacts within other ITTs. Figure 20 illustrates this point with a more complex ITT that interrelates two ITTs

Figure 19. Simple intertextual thread

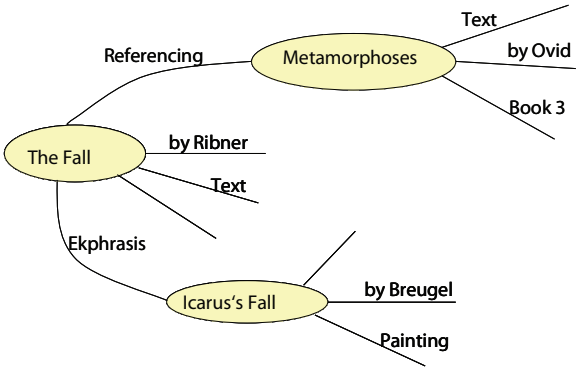
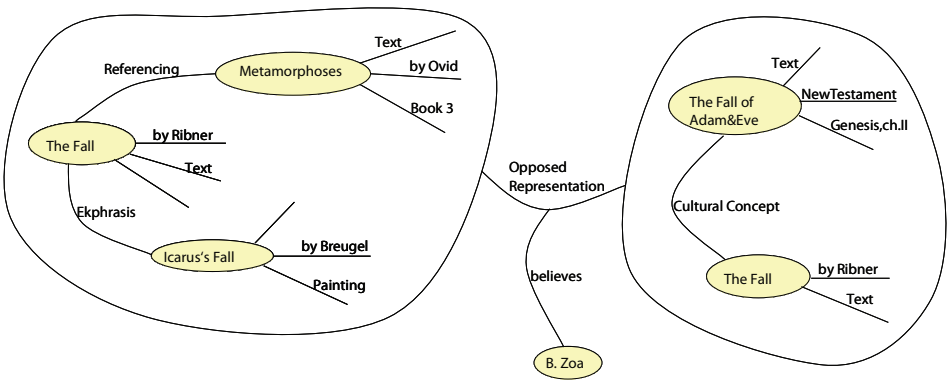


Figure 20. Complex intertextual thread



manifesting two different views on Ribner's poem as opposed representations.

Reification of statements is also frequently occurring within ITTs. Since experts in intertextual studies extensively base their position on the position of other researchers, statements about statements are common practice within ITTs. In the ITT of Figure 20, for instance, it is expressed by reification that the statement describing the two depicted ITTs as opposed representation is only the opinion of a certain researcher B. Zoa.

Given these characteristics of ITTs, we have found that Emmos are very well suited for their representation in the multimedia collaboration platform for intertextual studies that is envisioned by CULTOS. Firstly, the semantic aspect of Emmo offers sufficient expressiveness to capture ITTs. Figure 21 shows how the complex ITT of Figure 20 could be represented using Emmos. Due to the fact that associations as well as Emmos themselves are first-class entities, it is even possible to cope with reification of statements as well as with encapsulation of ITTs.

Secondly, the media aspect of Emmos allows researchers to enrich ITTs that so far expressed interrelationships between cultural artefacts on an abstract level with digital media about these artefacts, such as a JPEG image showing Breugel's

painting, Icarus' Fall. The ability to consume these media while browsing an ITT certainly enhances the comprehension of the ITT and the relationships described therein.

Thirdly, with the functional aspect of Emmos, functionality can be attached to ITTs. For instance, an Emmo representing an ITT in CULTOS offers operations to render itself in an HTML-based hypermedia view.

Additionally, our Emmo container infrastructure outlined in the previous section provides a suitable foundation for the realization of the CULTOS platform. Their ability to persistently store Emmos as well as their interfaces which enable applications to fine-grainedly traverse and manipulate the stored Emmos and invoke their operations make Emmo containers an ideal ground for the authoring and browsing applications for ITTs that had to be implemented in the CULTOS project. Figure 22 gives a screenshot of the authoring tool for ITTs that has been developed in the CULTOS project which runs on top of an Emmo container.

Moreover, their decentralized approach allows the setup of independent Emmo containers at the sites of different researchers; their ability to import and export Emmos with all the aspects they cover facilitates the exchange of ITTs, including the media by which they are enriched as well as the functionality they offer. This enables researchers to share and collaboratively work on ITTs in order to discover and establish new links between artworks as well as different personal and cultural viewpoints, thereby paving the way to novel insights to a subject. The profound versioning within the Emmo model further enhance this kind of collaboration, allowing researchers to concurrently create different versions of an ITT at different sites, to merge these versions, and to highlight differences between these versions.

Semantic Jukebox

One of the most prominent (albeit legally disputed) multimedia content sharing applications is the sharing of MP3 music files. Using peer-to-peer file sharing infrastructures such as Gnutella, many users gather large song libraries on their home PCs which they typically manage with one of the many jukebox programs available, such as Apple's

Figure 21. Emmo representing an ITT

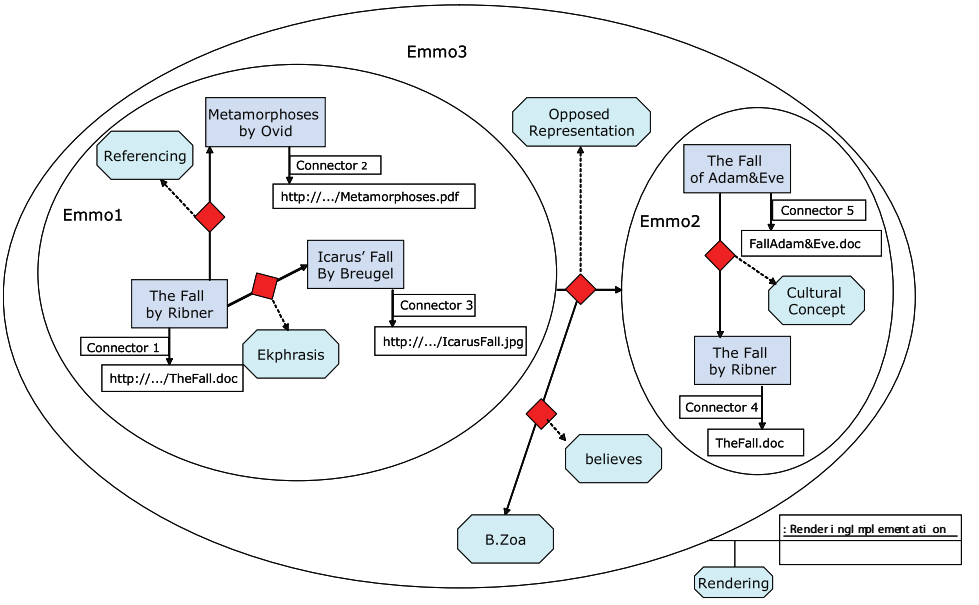
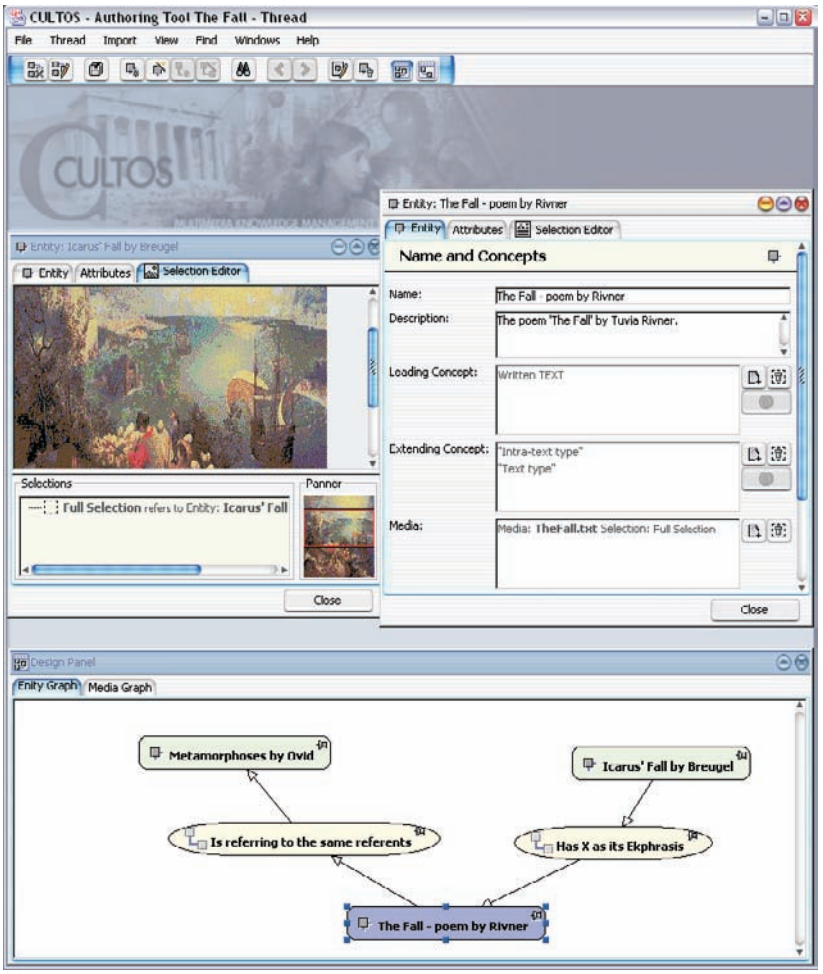


Figure 22. CULTOS authoring tool for ITTs

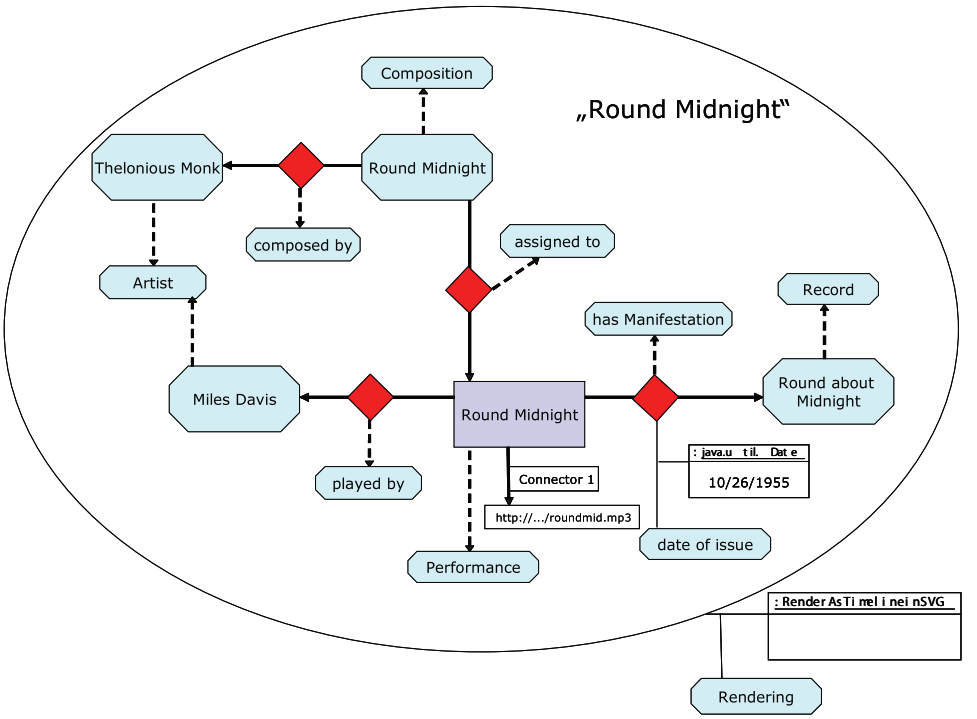


iTunes (Apple Computer, n.d.). The increasing use of ID3 tags (ID3v2, n.d.) — optional free text attributes capturing metadata like the interpreter, title, and the genre of a song - within MP3 files for song description alleviates the management of such libraries.

Nevertheless, ID3-based song management quickly reaches its limitations. While ID3 tags enable jukeboxes to offer reasonably effective search functionality for songs (provided the authors of ID3 descriptions spell the names of interprets, albums, and genres consistently), more advanced access paths to song libraries are difficult to realize. Apart from other songs of the same band or genre, for instance, it is difficult to find songs similar to the one that is currently playing. In this regard, it would also be interesting to be able to navigate to other bands in which artists of the current band played as well or with which the current band appeared on stage together. But such background knowledge cannot be captured with ID3 tags.

Using Emmos and the Emmo container infrastructure, we have implemented a prototype

Figure 23. Knowledge about the song "Round Midnight" represented by an Emmo



of a *semantic* jukebox that considers background knowledge about music. The experience we have gained from this prototype shows that the Emmo model is well-suited to represent knowledge-enriched pieces of music in a music sharing scenario. Figure 23 gives a sketch of such a music Emmo which holds some knowledge about the song "Round Midnight."

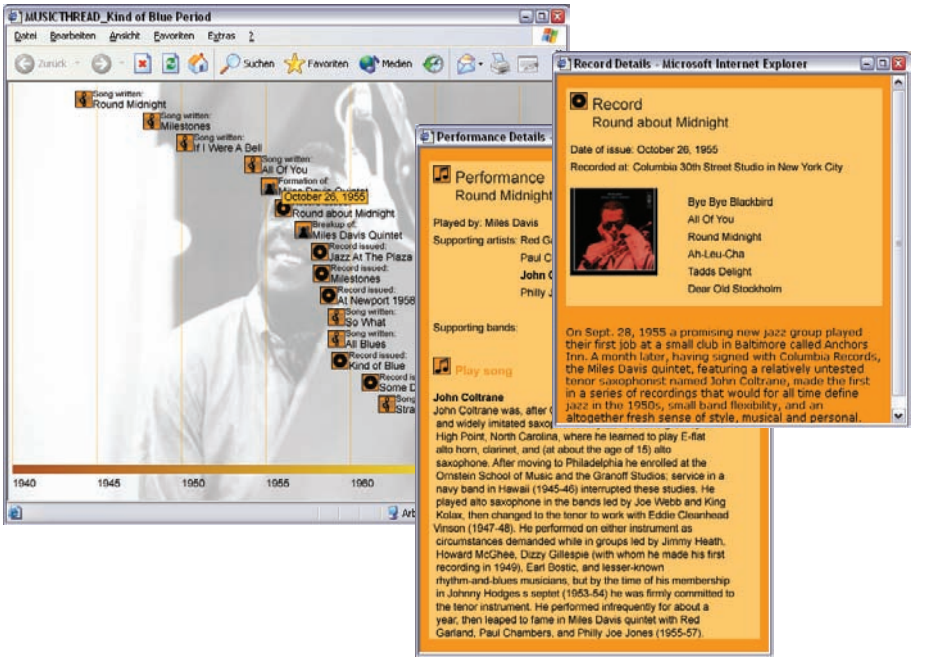
Its media aspect enables the depicted Emmo to act as a container of MP3 music files. In our example, this is a single MP3 file with the song "Round Midnight" that is connected as a media profile to the logical media part Round Midnight in the center of the figure.

The Emmo's semantic aspect allows us to express rich background knowledge about music files. For this purpose, we have developed a basic ontology for the music domain featuring concepts such as "Artist", "Performance", "Composition", and "Record" that all appear as ontology objects in the figure. The ontology also features various association types which allow us to express that "Round Midnight" was composed by Thelonious Monk and the particular performance by Miles Davis can be found on the record "Round about Midnight".

The ontology also defines attributes for expressing temporal information like the issue date of a record.

The functional aspect, finally, enables the Emmo to support different renditions of the knowledge it contains. To demonstrate this, we have realized an operation that, being passed a time interval as its parameter, produces an SVG timeline rendition (see screenshot of Figure 24) arranging important events like the foundation of

Figure 24. Timeline rendition of a music Emmo



bands, the birthdays and days of death of artists, and so forth, around a timeline. More detailed information for each event can be gained by clicking on the particular icons on the timeline.

Further operations could be imagined; for example, operations that provide rights clearance functionality for the music files contained in the Emmo, which is a crucial issue in music sharing scenarios.

Our Emmo container infrastructure provides a capable storage foundation for semantic jukeboxes. Their ability to fine-grainedly manage Emmos as well as their scalability allowing them to be deployed as both small-scale file-based and as large-scale database server configurations. Thus, Emmo containers constitute suitable hosts for knowledge-enriched music libraries of private users as well as libraries of professional institutions such as radio stations. Capable of exporting and importing Emmos to and from bundles, Emmo containers also facilitate the sharing of music between different jukeboxes. Their versioning support even allows it to move from mere content sharing scenarios to collaborative scenarios where different users cooperate to enrich and edit Emmos with their knowledge about music.

CONCLUSION

Current approaches to semantic multimedia content modeling typically regard the basic media which the content comprises, the description of these media, and the functionality of the content as conceptually separate entities. This leads to difficulties with multimedia content sharing and collaborative applications. In reply to these difficulties, we have proposed Enhanced Multimedia Meta Objects (Emmos) as a novel approach to semantic multimedia content modeling. Emmos

coalesce the media of which multimedia content consists, their semantic descriptions, as well as functionality of the content into single indivisible objects. Emmos in their entirety are serializable and versionable, making them a suitable foundation for multimedia content sharing and collaborative applications. We have outlined a distributed container infrastructure for the persistent storage and exchange of Emmos. We have illustrated how Emmos and the container infrastructure were successfully applied for the sharing and collaborative authoring of multimedia-enhanced intertextual threads in the CULTOS project and for the realization of a semantic jukebox.

We strive to extend the technological basis of Emmos. We are currently developing a query algebra, which permits declarative querying of all the aspects of multimedia content captured by Emmos, and integrating this algebra within our Emmo container implementation. Furthermore, we are wrapping the Emmo containers as services in a peer-to-peer network in order to provide seamless search for and exchange of Emmos in a distributed scenario. We also plan to develop a language for the definition of ontologies that is adequate for use with Emmos. Finally, we are exploring the handling of copyright and security within the Emmo model. This is certainly necessary as Emmos might not just contain copyrighted media material but also carry executable code with them.

REFERENCES

Apple Computer (n.d.). *iTunes.* Retrieved 2004 from *http://www.apple.com*

Ayars, J., Bulterman, D., Cohen, A., et al. (2001). *Synchronized multimedia integration language (SMIL 2.0).* W3C Recommendation, World Wide Web Consortium (W3C).

Baumeister, S. (2002). *Enterprise media beans TM specification*. Public Draft Version1.0, IBM Corporation.

Benari, M., Ben-Porat, Z., Behrendt, W., Reich, S., Schellner, K., & Stoye, S. (2002). Organizing the knowledge of arts and experts for hypermedia presentation. *Proceedings of the Conference of Electronic Imaging and the Visual Arts,* Florence, Italy.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The semantic web.* Scientific American.

Boll, S., Klas, W., & Westermann, U. (2000). Multimedia document formats - sealed fate or setting out for new shores? *Multimedia - Tools and Applications, 11*(3).

Brickley, D., & Guha, R.V. (2002). *Resource description framework (RDF) vocabulary description language 1.0: RDF Schema*. W3C Working Draft, World Wide Web Consortium (W3C).

Chang, H., Hou, T., Hsu, A., & Chang, S. (1995). Tele-Action objects for an active multimedia system. *Proceedings of the International Conference on Multimedia Computing and Systems (ICMCS 1995),* Ottawa, Canada.

Chang, S., & Znati, T. (2001). Adlet: An active document abstraction for multimedia information fusion. *IEEE Transactions on Knowledge and Data Engineering, 13*(1).

Daniel, R., Lagoze, D., & Payette, S. (1998). A metadata architecture for digital libraries. *Proceedings of the Advances in Digital Libraries Conference,* Santa Barbara, California.

Fensel, D. (2001). *Ontologies: A silver bullet for knowledge management and electronic commerce*. Heidelberg: Springer.

Ferraiolo, J., Jun, F., & Jackson, D. (2003). *Scalable vector graphics (SVG) 1.1.* W3C Recommendation, World Wide Web Consortium (W3C).

Gnutella (n.d.). Retrieved 2003 from *http://www.gnutella.com*

Grimson, J., Stephens, G., Jung, B., et al. (2001). Sharing health-care records over the internet. *IEEE Internet Computing, 5*(3).

ID3v2 (n.d.). [Computer software]. Retrieved 2004 from *http://www.id3.org*

ISO/IEC JTC 1/SC 29 (1997). *Information technology - Coding of hypermedia information - part 5: support for base-level interactive applications.* ISO/IEC International Standard 13522-5:1997, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC).

ISO/IEC JTC 1/SC 29/WG 11 (2001). *Information technology - Multimedia content description interface - part 5: Multimedia description schemes.* ISO/IEC Final Draft International Standard 15938-5:2001, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC).

ISO/IEC JTC 1/SC 34/WG 3 (1997). *Information technology - Hypermedia/time-based structuring language (HyTime).* ISO/IEC International Standard 15938-5:2001, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC).

ISO/IEC JTC 1/SC 34/WG 3 (2000). *Information technology - SGML applications - topic maps.* ISO/IEC International Standard 13250:2000, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC).

ISO/JTC1/SC 32/WG 2 (2001). *Conceptual graphs.* ISO/IEC International Standard, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC).

Lagoze, C., Lynch, C., & Daniel, R. (1996). *The warwick framework: A container architecture for aggregating sets of metadata*. Technical Report TR 96-1593, Cornell University, Ithaca, New York.

Lassila, O., & Swick, R.R. (1999). *Resource description framework (RDF) model and syntax specification.* W3C Recommendation, World Wide Web Consortium (W3C).

Leach, P. J. (1998, February). *UUIDs and GUIDs.* Network Working Group Internet-Draft, The Internet Engineering Task Force (IETF).

Matena, V., & Hapner, M. (1998). *Enterprise Java Beans* TM. Specification Version 1.0, Sun Microsystems Inc.

Nejdl, W., Wolf, B., Qu, C., et al. (2002). EDUTELLA: A P2P networking infrastructure based on RDF. *Proceedings of the Eleventh International World Wide Web Conference (WWW 2002)*, Honolulu, Hawaii.

Newmann, D., Patterson, A., & Schmitz, P. (2002). *XHTML+SMIL profile.* W3C Note, World Wide Web Consortium (W3C).

Pereira, F., & Ebrahimi T., (Eds.) (2002). *The MPEG-4 book*. CA: Pearson Education

Reich, S., Behrendt, W., & Eichinger, C. (2000). Document models for navigating digital libraries. *Proceedings of the Kyoto International Conference on Digital Libraries*, Orlando, Kyoto, Japan.

Raggett, D., Le Hors, A., & Jacobs, I. (1999). *HTML 4.01 specification.* W3C Recommendation, World Wide Web Consortium (W3C).

ENDNOTE

[1] See http://www.cultos.org for more details on the project.

This work was previously published in Managing Multimedia Semantics, edited byU. Srinivasan & S. Nepal, pp. 305-332, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 2.4
Designing for Learning in Narrative Multimedia Environments

Lisa Gjedde
Danish University of Education, Denmark

ABSTRACT

Narrative is fundamental for learning and the construction of meaning. In the design of interactive learning programs, the need for narrative is often neglected, and the emphasis is on information design rather than the design of experiential learning environments. This chapter presents research related to the development of two prototypes of narrative interactive multimedia learning environments, from an experiential and situated learning perspective and proposes a model for a narrative learning process, related to a situated and experiential learning perspective.

BACKGROUND

Narrative is fundamental for the construction of meaning on a personal as well as on a community level. The narrative format has been a traditional way of teaching in many cultures, and teachers may develop a competence as storytellers, drawing on narrative for motivation, and for experiential and contextual learning by using stories or having the learner's develop stories themselves (Gudmundsdottir, 1991). The concept of narrative encompasses both the narrative expression in the form of story-making and narrative as a cognitive tool for the construction of knowledge, which includes the construction of culturally embedded

knowledge as well as being an important part of knowledge sharing (Bruner, 1990; Schank, 1995). It also plays an important part in collaborative and experiential learning. The concept of situated and collaborative learning has been put forth by Brown, Collins, and Duguid, (1989) in their seminal article on Situated Cognition and the Culture of Learning: "Learning, both outside and inside school, advances through collaborative social interaction and the social construction of knowledge." The role of narrative in this learning process is important for distributed and embedded knowledge.

NARRATIVE AND MULTIMEDIA

Research into interactive multimedia as a resource for instruction and learning has previously pointed to problems in the organization and presentation of the material in relation to the cognitive processes of the students. Researchers from the MENO-project (Multimedia in Education and Narrative Organisation), located at the Open University and University of Sussex, have been investigating the role of narrative in relation to comprehension and learning in interactive multimedia, based on findings that the degree of narrative structure would affect the learners' level of comprehension. They found (Laurillard et al., 1998) that "learners working on interactive media with no clear narrative structure display learning behaviour that is generally unfocused and inconclusive." Based on a hypothesis of narrative as fundamental for learning, they designed an experimental study with three versions of material on a CD-ROM, with different degrees of narrative structure, and tested the different versions in classroom settings. The CD-ROM offered video sequences and a notepad for collecting material as well as questions to guide the exploration. Their conclusions point to the importance of designing interactive multimedia environments (Plowman et al., 1999), "so learners are able to both find narrative coherence and generate it for themselves." Groundbreaking cognitive psychologist and AI-researcher Roger Schank (1995), has likewise found narrative reasoning and construction to be fundamental for cognitive processes. At a theoretical level, Schank (with Abelsson) has contributed seminally through his cognitive models of goals, scripts, and plans. At the level of development and research, he has developed several prototypes and multimedia training programs that draw on narrative elements like case stories. He has advocated different learning architectures, which offers the possibility for sharing knowledge and natural learning and that stories can be used as a fundamental mode of communication and learning not only in case-based learning architectures but also in exploratory and incidental learning. Schank's prototypes of learning architectures with their call for life-like learning situations can be seen as exponents for experiential learning (Schank & Cleary, 1995).

The above-mentioned projects focused on their exploration of narrative in relation to educational and cognitive processes. They, however, do not provide prototypes that afford an exploration of the potentials of learning in an open environment, where the learner's own production of multimedia narratives is supported by offering multimedia tools.

Experiential learning is most often characterized by learning from primary experience. In relation to the research projects presented here, it is suggested that a fictional and imagination-based experience may offer a context for experiential learning that may provide material for expression and reflection. An influential model for experiential learning, which has been developed by David A. Kolb, draws on the work of John Dewey and Jean Piaget. However, in relation to the development of principles for the design of narrative interactive learning environments, it is important to include a focus on the situated and social aspects of the learning process.

There are potential learning advantages to be gained by constructing multimedia learning environments, not primarily as way of presenting encyclopedia-style factual material, but rather through offering the learners an experiential pathway into the material. It is further suggested that this may be enhanced by providing a narrative experience through a narrative framework, through narrative visual elements, as well as offering multimedia tools for narrative construction.

This chapter raises some questions related to the development of such prototypes of narrative interactive multimedia learning environments from an experiential and situated learning perspective and suggests a model for a narrative learning process, related to a situated and experiential learning perspective.

DESIGN OF NARRATIVE INTERACTIVE LEARNING ENVIRONMENTS

Designers of educational interactive multimedia programs are facing a challenge as to how to design learning environments that may enhance the collaborative learning that is situated as well as cognitive, through the use of narrative in the design and content. They are also facing the challenge of how to engage the learner in interactive experiences that are meaningful at the level of participation and expression of the learner. The participation may be in an immersive experience that allows the learner to actively engage in the action or to identify with the protagonist of the story. It may also be an experience with elements of role-playing, where the learner takes the action forward by using imagination as well as factual knowledge, and thus interacts by moving the story line forward. Engaging the learner in an imaginative productive way, where the learner is guided to participate in the development of the narrative plot and contribute to the development of the story, may be a way of engaging the learner at multiple levels of involvement.

Often, the use of story in education presents the teacher in a role as storyteller and the students as listeners that follow the unfolding of the narrative. The use of narrative in educational media calls for ways to engage the learner in a process of narrative that is meaningful and fully draws on the possibilities of the media for creation of media expression by the user, and thus for engaging the learner at different levels.

By placing the learner in the role as a storyteller, the learning context becomes one that may enhance the level of involvement as well as the inclusiveness of the situation, as the learner may approach the task from a different point of view and use different modalities of expression: images, sound, movement, etc.

Another issue to be explored in the development of narrative multimedia learning environments concerns the creation of content that is meaningful, that may be engaging for the learner and takes into consideration possible differences in gender-specific interests and social and cultural backgrounds.

Some of these issues related to the development of an architecture and design for narrative based on principles for experiential and situated learning, with the creation of authentic tasks, have been explored in two research projects funded by the Danish Ministry of Education, as part of a program to further the development and research of the use of ICT in schools.

PARTICIPATORY CONTENT DESIGN

The Narrative Universes of Children

This project is on the creation of content and the design of interactive multimedia narratives by children within the Danish Language and Art curriculum. The project involved students of Grades 4 to 6 in a study drawing on principles of participatory design. Four classes at four different schools, numbering about 100 students and

six teachers, participated in this project, which aimed at exploring children's narrative universes and their narrative construction and collaboration in multimedia productions and the creation of content design supporting that.

The research design included two separate phases. The first phase was an investigation of the narrative universes, which the children in this age group produced, when prompted to produce stories and develop imagery to illustrate them. The children produced stories and drawings, working in groups, for a period of weeks. They were allowed to use different modes of expression, ranging from making giant storybooks to be used for the kindergarten pupils, to the use of tape recorders and puppet theatre. The main purpose with this first phase was to prompt the creation of narrative material in order to inform the design process for the animation program. In this way, the learners would help create relevant content material, which subsequently could be used in relation to the animation program they were to use in the second phase.

The second phase focused on the learner's construction of interactive narratives using a simple animation program that allowed for production of interactive multimedia narratives by the learners. Based on the analysis of the stories and narrative universes created by the children during the first phase, a series of narrative universes was developed by the researchers based on the content created by the children. These universes were then produced into graphic material, by professional artists, in order to be used by the learners in the animation program.

The animation program allowed the learner to use these images and also to construct and modify the images and animations and the use of text and sound. This led to the creation of a rich material of interactive stories provided by the learners as well as classroom observation of their collaborative processes in their construction of the interactive narratives. This process in which most of the students where highly motivated, led to other processes of informal collaborative learning leading to mastery of the program and production of the story in which they were involved.

The collaborative learning was expressed in the creation of shared environments, with narrative elements, e.g., characters, settings, and themes. The program aimed at creating a context that would support the development of the learners' narrative competences and their construction of interactive narratives that were relevant to them, and it would also support their sharing of experiences. A further analysis of this rich material may reveal more of the learner's strategies for developing interactive media literacy and expressive nonlinear narrative competencies, and how this type of learning environment may support situated and collaborative learning and knowledge construction.

NARRATIVE FRAMEWORK – NARRATIVE EXPRESSION

Another project related to the development and research of narrative multimedia learning environments is the project "Narrative in interactive Web-based learning environments." This project explores the interplay between the narrative context that includes a narrative setting, with access to resources including stills, audio, and text, and the learner's collaborative production of narratives using these resources. The prototypes are produced and explored as action research using a qualitative methodology. It is an iterative design process involving an educational setting with three secondary schools with students in Grades 8 and 9.

In order to facilitate this process of collaborative learning, it is important that the context be motivating and appealing and that it offer familiar and interesting figures, settings, and issues that may be the frame for further exploration and narrative development relating to that theme (Gjedde, 2002). Part of the development

Figure 1. Multimedia narrative interface

of the script for the frame narrative included the showing of documentary film material from the historical period that this multimedia learning program addressed. This film was shown to the target group of Grade 8 students, and they were asked to fill out a questionnaire, subsequently, that focused on the different content areas of the film about the German occupation of Denmark during the Second World War, rating it for interest. The questionnaire and discussion of the film showed gender-specific preferences for the subjects. These were addressed in the development of the script for the frame narrative in order to provide content material, which would be interesting and motivating for the different students in relation to their preferences.

The illustrations that are used in this Web-based interactive multimedia program were made by an award-winning Danish artist. By applying a certain aesthetic expression on purpose, they are meant to address the target group and help create an immersive experience.

Figure 1 shows part of the room, which is the main interface in the program. It is navigable at 360 degrees, allowing the learner to explore the room and the artifacts, which are related to the narrative learning process.

Most of the artifacts are interactive and will open up to the activities that are available in this learning environment. The activation of the picture over the desk, which is indicated on the figure, will take the learner into the frame narrative. The frame narrative will guide the learner through some fictional but historically correct episodes, and lead the learner to some questions to continue the storytelling, either from the point of view of the male or female protagonist. The book on the table will provide access to a specially designed Web builder. Through this production tool, material from the databases holding sound and image files can be integrated in the Web-based multimedia stories produced by the learners. All the relevant source materials and production tools can be accessed as artifacts in the room, being held, for instance, in the cupboard, the book on the table, and the phone. This coherence between the graphic interface and the narrative environment is done in order to provide for a *pervasive narrative experience*, in which all actions and all materials are embedded in the context of the frame narrative. This is done in order to support the use of a narrative logic and narrative reasoning involved in the narrative experience, which provides for a shared learning environment that may facilitate

the building of stories involving shared knowledge and experience in a community of practice.

TOWARD A NARRATIVE LEARNING PROCESS

Storytelling and the use of narrative in interactive multimedia programs can offer ways of engaging with the material at different levels. The narrative elements and the way they are embedded in an interactive learning environment can serve as cognitive artifacts, which support the construction of meaning through structure, content, and culturally implicated values. The narratives offer immersive experiences, which allow the user to engage at an emotional level and involve the user with different emotional states.

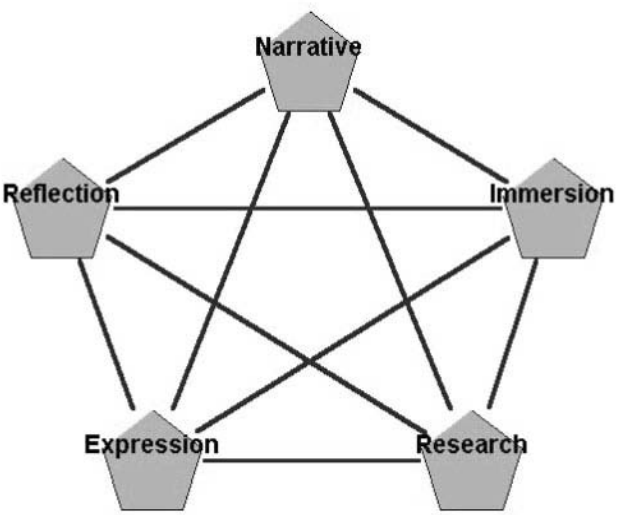
Developing and producing stories stimulates narrative thinking and a sense of causal relationships and analogy. It brings about possibilities for identification with the characters and situations, and in this way, it allows for multiple perspectives on situations and different points of view that are essential social skills. The development of this level of understanding and the competence to interpret and construct meaning are important skills and values in terms of the personal development and sense of cultural identity and personal values of the learner.

The stories offer materials for interpretation and reflection, and a shared cultural environment for doing so. In the interactive learning environment, the reflective process is distributed by the use of interactive log books and online conferences that allows for reflection on the process and productions.

The narrative learning process is framed by stories that are contextual, that make up the shared conceptual learning environment, which is characterized by its setting, characters, actions, and themes. The learner's development of the stories may lead to a process of research into the elements and factual material that are necessary for the unfolding of the story. The process of the articulation of the narratives in an interactive multimedia production may lead to the sharing of knowledge in a way that is socially and cognitively inclusive.

Through the narrative expression and articulation in digital media, a process that is experiential as well as situated in communities of practice, may

Figure 2. Narrative learning process



be initiated. This model on a narrative learning process is presented in Figure 2.

The model for the narrative learning process in the interactive multimedia learning environment described above, involves five distinct learning approaches and activities that are interrelated and mutually supportive. This model suggests a process of dynamic learning in which the elements are synergically related:

1. *Narrative*. The top point of the pentad refers to the concept of narrative that is based on the premise that narrative is an important experiential and contextual frame and approach to learning. It relates to the narrative content as well as to the organizing narrative structure.
2. *Immersion*. The concept of immersion is related to the experience of narrative content and structure. It is through the story and the related imaginative story elements that the learner experiences the immersion. It may be experienced further in the learning process as immersion in the subject matter and a research activity that is then expressed in the learner's own narrative articulation.
3. *Research*. The activity of research is a learner-directed activity delving into information and themes related to the narrative and the curricular areas it is exploring. It is focused by the narrative and supported by the sense of immersion into the subject.
4. *Expression*. The activity of expression is based on the research activity and is held in line by the narrative, which is the context it relates to and it further elaborates on or uses as a base for the construct of new narratives. It is also supported by the sense of immersion into the narrative context and characters.
5. *Reflection*. This activity can be a reflection by the learners on the material presented in narrative or nonfiction form, as well as on the narrative expression produced by the fellow learners. It is thus offering the teacher an avenue for reflecting on the learners' concepts as they are expressed in narrative form. It can encompass a hermeneutic as well as an aesthetic, imaginative, and expressive process.

These above-mentioned points are relevant parameters to include in the evaluation of the learning potentials of narrative interactive learning environments. Key questions to be asked in such an evaluation include the following:

- To what extent does the narrative interactive learning environment offer a narrative experience with a content that is relevant, engaging, and pervasive?
- Does it provide tools for the learner to participate in a process of articulation?
- Is it to be used with a learning scenario that includes a process of reflection?
- What roles does it offer the learner and the teacher?

CONCLUSION

Experiential and narrative learning processes move the focus from the instructor toward the learner. It moves it from the transfer of information, toward the hermeneutic process of interpretation and construction of meaning. By building architectures for learning in interactive multimedia environments that take into account the hermeneutic narrative process and the expressive narrative process, reflection on the learner's own expression is included.

Principles for the development of narrative multimedia learning environments must ideally be based on educational and psychological principles as well as on including production theory, in order to offer the potentials for a satisfactory learning experience.

Using the approach of narrative learning may further the knowledge of how to design and make available an interactive multimedia learning environment that can support the learner toward the understanding of the events and the meaning they hold and not just aim at achieving knowledge of factual events. Having this as the overarching learning paradigm supporting the design, it may allow for the learners to enter into levels of learning in which they may be more likely to be involved at personal levels. Thus, the boundaries between formal and nonformal learning environments may be blurred and allow for the learners to be part of communities of practice that are situated in experiential settings and cater to their needs to express their own goals and learning.

ACKNOWLEDGMENT

The projects "The Narrative Universes of Children" and "Narrative in interactive Web-based learning environments" are funded by the Danish Ministry of Education, "ITMF-programme."

The project "The Narrative Universes of Children" has been carried out in collaboration with Leif Gredsted, Associate Professor, DPU.

REFERENCES

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

Dewey, J. (1939). *Education and experience*. New York: Collier Books.

Gjedde, L. (2002). Context, cognition and narrative experience in Sophies World. In B. H. Sørensen, & O. Danielsen (Eds.), *Learning and narrativity in digital media*. Samfundslitteratur: København.

Gudmundsdottir, S. (1991). Story-maker, storyteller; narrative structures in curriculum. *Journal of Curriculum Studies*, *23*(3), 207–218.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice Hall.

Laurillard, D., Stratfold, M., Luckin, R., Plowman, L., & Taylor, J. (1998). Multimedia and the learner's experience of narrative. *Computers and Education*, *31*, 229–242.

Mandler, J. (1984). *Stories, scripts, and scenes: Aspects of Schema Theory.* Hillsdale, NJ: Erlbaum.

Plowman, L., Luckin, R., Laurillard, D., Stratfold, M., & Taylor, J. (1999). Designing multimedia for learning: Narrative guidance and narrative construction. *Proceedings CHI'99:ACM Conference on Human Factors in Computing Systems* (pp. 310–317). Pittsburgh, PA, USA,15–20 May.

Schank, R. C., & Abelson, R. P. (1995). *Knowledge and memory: The real story.* J. R. S. Wyer. Hilldale, NJ: LEA.

Schank, R. C., & Cleary, C. (1995). *Engines for education*. Hilldale, NJ: LEA.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra & R.C. Sharma, pp. 101-112, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.5
Multimedia Learning Designs:
Using Authentic Learning Interactions in Medicine, Dentistry, and Health Sciences

Mike Keppell
Hong Kong Institute of Education, Hong Kong

Ngaire Kerse
University of Auckland, New Zealand

Jane Gunn
The University of Melbourne, Australia

Karen Kan
The University of Melbourne, Australia

Kelsey Hegarty
The University of Melbourne, Australia

Louise Brearley Messer
The University of Melbourne, Australia

Vivienne O'Connor
The University of Queensland, Australia

Heather Bione
The University of Melbourne, Australia

ABSTRACT

This chapter describes the learning design of two multimedia modules which complement a problem-based learning health sciences curriculum. The use of student-centred, authentic learning design frameworks guide academics and instructional designers in the creative pedagogical design of learning resources. The chapter describes the educational context, learning design of two multimedia modules and suggests a number of strategies for improving the design and development of multimedia resources.

INTRODUCTION

This chapter examines the instructional design of two multimedia modules that utilize authentic learning interactions to teach medical, dental, and health science concepts. Interactive multimedia modules complement the broader goals of a problem-based learning curriculum and enrich the health science curriculum by addressing conceptually difficult content areas. It is essential that the learning design (Koschmann, Kelson, Feltovich, & Barrows, 1996) of self-directed learning modules "should be informed from its

inception by some model of learning and instruction" (p. 83). The use of student-centered learning approaches is becoming increasingly popular in medicine, dentistry, and health science curricula as the teaching of problem-based learning and case-based learning assure a close match with real-world clinical cases. This chapter outlines the educational context and then examines two multimedia modules that utilize a case-based learning design. As educators, it is essential that we articulate our learning design for educational interventions from the earliest stages so that we are able to integrate the module into the educational setting and also provide a framework for evaluating the innovation (Koschmann, Kelson, Feltovich, & Barrows, 1996).

EDUCATIONAL CONTEXT

The medical course at the University of Melbourne had traditionally been taught using a discipline-based approach. Internal review mechanisms and student feedback in recent years had highlighted a number of deficiencies in the traditional course. In broad terms, these included insufficient integration between the basic and clinical sciences, insufficient attention to teaching communication skills, problem-solving skills, and social aspects of health, and an overload of biomedical detail that was duplicated in subjects originating from different departments. In an effort to remedy these deficiencies and also to incorporate current theories of medical education, a new medical curriculum was introduced in 1999. The pedagogical model for the new medical curriculum incorporates elements of problem-based learning (PBL) and self-directed learning (SDL) (Koschmann, Kelson, Feltovich, & Barrows, 1996). The primary focus of learning in semesters 2 through 5 is through medical problems (known as problems of the week), which are presented to students in small group tutorial settings. A key feature of the new curriculum is the horizontal integration across disciplines and the vertical integration of clinical situations with basic scientific material (Keppell, Kennedy, Elliott, & Harris, 2001).

The transformation of a medical course by the faculty involved considerable analysis, planning, investment of resources, staffing, and fundamental changes to teaching and learning approaches. Changing a traditional teaching and learning approach to a PBL approach represented a major pedagogical shift for academics within the faculty. In addition to this fundamental change, the curriculum also placed considerable emphasis on Web-based and multimedia teaching resources. This major departure from a traditional medical curriculum required support from academic staff across a wide range of diverse disciplines. A number of researchers have examined the importance of context in supporting major change like the faculty's curricular transformation. According to Altschuld and Witkin (2000), the following was found:

> *…implementation of innovations required awareness on the part of and strong support from administrators, a dedicated and critical mass of staff working with the change, communication channels that were frequently used by that staff to promote change, and a climate that makes nonadopters feel as though they are "out of it" unless they begin to adopt or move forward* (p. 182)

These factors will make or break the implementation of the innovation.

Use of Information and Communication Technology in the Curriculum

The use of information and communication technology (ICT) is an important feature of the problem-based learning curriculum. ICT is utilized to deliver medical content in two ways.

These include the use of the Web-based problem of the week embedded in the TopClass learning framework and stand-alone, computer-facilitated learning modules. The TopClass learning framework provides a central access point for students to enter the online course work, complete self-assessment tests, view class announcements, participate in discussion groups, and send and receive messages from teachers or peers. In the first three years of the medical curriculum, the medical student needs to complete 60 problems of the week. Approximately 120 problems of the week will be required for the entire curriculum. Self-directed learning resources, in the form of computer-facilitated learning modules, are also required to support the core content of the problems. Approximately 70 modules are in use or are currently under development within the faculty, and it is envisaged that 100 modules will be required to support the curriculum in its entirety.

Self-directed learning resources have a unique role in the curriculum. These resources are often initiated by clinicians who teach conceptually difficult content areas. A conceptually difficult content area (usually one to two hours) is often identified and developed by academics and multimedia teams. For instance, a module on pediatric dentistry focusing on a diabetic child complements a dental curriculum; a module on sensitive examination technique (SET) cervical screening is used to complement the teaching of cervical cancer. These resources may be used independently of the PBL curricula or they may be developed to complement the content within a problem of the week. Often, these computer-facilitated learning modules support the core content of the problems. Figure 1 portrays the use of ICT in the curriculum and the use of computer technology to complement, enhance, and support teaching and learning in the curriculum. The focus of this chapter is on the self-directed learning component of Figure 1.

MULTIMEDIA AND ONLINE MODULES

Figure 2 shows the variety of modules developed by the faculty as of December 2002. Each content

Figure 1. Use of computer technology in the medical curriculum

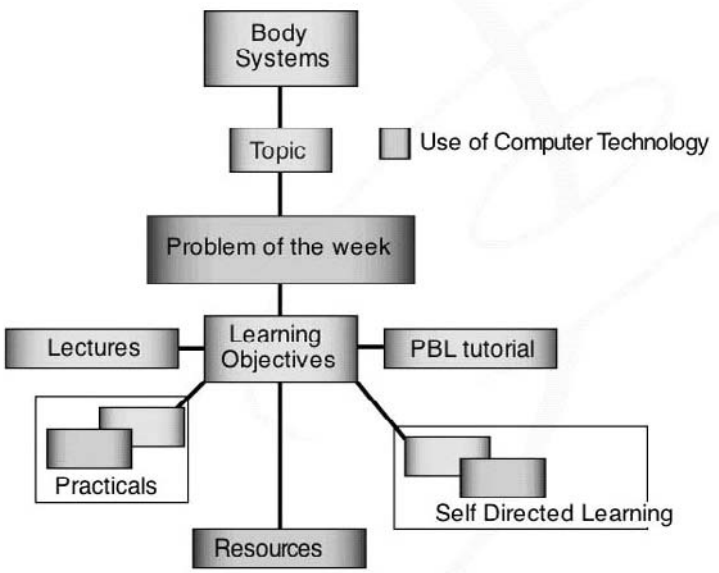
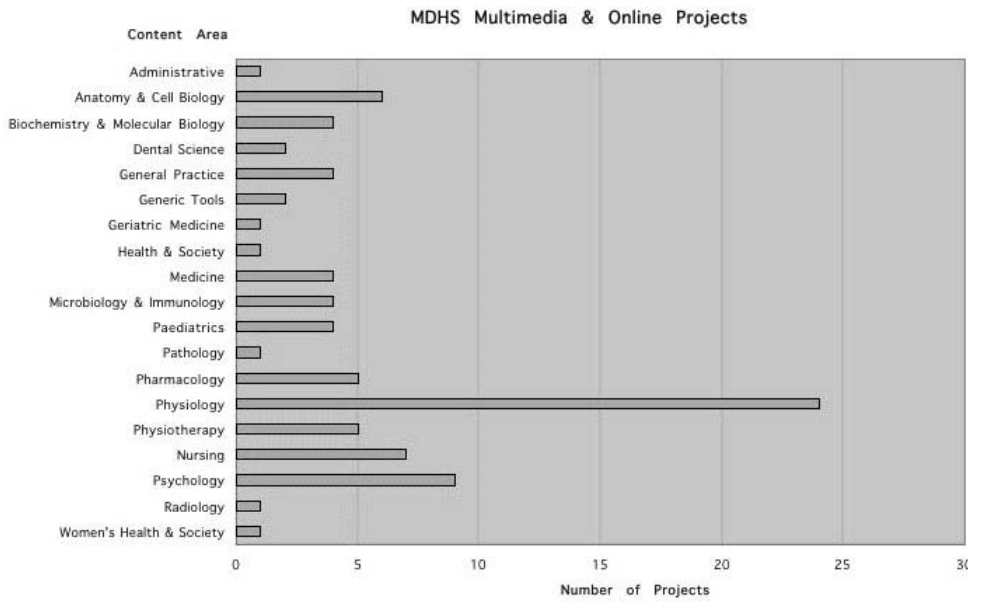


Figure 2. Multimedia and online modules utilized within the faculty



area is outlined with the number of modules per discipline area. This graphic provides an overview of the diversity of content areas that utilize multimedia to enhance teaching and learning. Many modules utilize student-centered teaching and learning approaches. In some instances, instructional designers have worked with the academics to develop their module utilizing constructivist teaching and learning principles, including case-based learning. The following aspects of this chapter examine the application of these principles to the design of two multimedia modules.

STUDENT-CENTERED LEARNING

There has been a trend away from teacher-directed instructional approaches to student-centered learning environments. Jonassen and Land (2000) compare the two methods in Table 1. In particular, in student-centered learning environments, contextualized, authentic, and situated learning interactions are emphasized. These principles have been adopted in the learning design of the two specific modules that will be discussed in this chapter, which utilize authentic learning interactions.

LEARNING DESIGN

Although, as stated in Herrington, Oliver, and Reeves (2002), "it is impossible to design truly authentic learning experiences" (p. 60), we attempted to develop learning experiences that would complement and enhance the professional practice of doctors and dentists. We considered a range of factors including the intent of the curriculum, lecturer teaching style, and learning outcomes in designing the multimedia modules. Authentic learning experiences, according to Jonassen, Mayes, and McAleese (1992), are "those which are problem- or case-based, that immerse the learner in the situation requiring him or her to acquire skills or knowledge in order to solve the problem or manipulate the situation" (p. 235).

Table 1. A comparision of instructive and student-centered learning environments

Instruction	Student-Centered Learning Environments
Transmission, acquisition	Interpretation, construction
Mastery, performance	Meaning making
External reality	Internal reality
Abstract, symbolic	Contextualized, authentic, experiential
Individually interpreted	Socially negotiated, coconstructed
Individual	Collaborative
Encoding, retention, retrieval	Articulation and reflection
Symbolic reasoning	Situated learning
Psychology	A nthropology, sociology, ethnography
Laboratory	in situ
Well-structured	Ill-structured
Decontexualized	Embedded in experience

Source: Adapted from Jonassen and Land (2000)

And, according to Young (1993), "Authentic tasks enable students to immerse themselves in the culture of the academic domain, much like an apprentice" (p. 43). Authentic learning contexts such as the virtual dental clinic and cervical screening module may have a number of advantages over more decontextualized teaching and learning settings. The authentic nature of the technology-enhanced, student-centered learning environment may anchor knowledge in authentic contexts. An effective learning environment enables learners to use its resources and tools to process more deeply and extend thinking (Jonassen, 1996; Jonassen & Reeves, 1996; Kozma, 1987).

In order to create realistic learning experiences for the students, it was essential that we immerse the student in authentic cases to guide our creation of the two modules. Herrington, Oliver, and Reeves (2002) outlined 10 characteristics of authentic activities. These comprise the following:

- Authentic activities have real-world relevance.

- Authentic activities are ill-defined, requiring students to define the tasks and subtasks needed to complete the activity.
- Authentic activities comprise complex tasks to be investigated by students over a sustained period of time.
- Authentic activities provide the opportunity for students to examine the task from different perspectives, using a variety of resources.
- Authentic activities provide the opportunity to collaborate.
- Authentic activities provide the opportunity to reflect.
- Authentic activities can be integrated and applied across different subject areas and lead beyond domain-specific outcomes.
- Authentic activities are seamlessly integrated with assessment.
- Authentic activities create polished products valuable in their own right rather than as preparation for something else.
- Authentic activities allow competing solutions and diversity of outcome.

Two modules will be analyzed to demonstrate how the above principles have been utilized to design authentic learning interactions. In the first instance, Tables 2 and 3 outline an overview of the principles and their concrete applications.

Authentic Activities have Real-World Relevance

In designing the two modules, we considered the "real-world tasks of professionals" (Herrington, Oliver, & Reeves, 2002, p. 62) as a basis for the module. For instance, in the field of pediatric dentistry, there are concerns that dental students are not competent in combining preventive and restorative management philosophies while integrating diagnosis and treatment planning (Suivinen, Messer, & Franco, 1998). Declining patient numbers and a need to focus on integration suggested that the use of multimedia case simulations were a viable alternative, as they replicate the dental clinic without requiring live patients. A module on diabetes and its implications for dentistry provides an opportunity to develop and consolidate the concept of integrated patient care. The diabetes case created unique challenges for the design team. A virtual pediatric diabetic patient was created in order to address the difficulty of obtaining relevant patient photographs. Our explicit learning design focused on increasing the level of student engagement with the content by contextualizing the scenario content within a virtual dental clinic (Keppell, Kan, Messer, & Bione, 2002).

The SET project attempts to examine a real-world clinical case of cervical screening so that medical students can decide which women should be screened on the basis of evidence-based screening recommendations. It also discusses the barriers to cervical screening from the patient's and doctor's perspectives and effective communication skills to explain how a Pap test is performed. Videos are sequenced to explain and demonstrate the steps involved in taking a Pap test and how to communicate the results to a woman. Currently, medical students have few opportunities to observe or perform Pap tests, and current literature documents negative screening experiences by women as a major barrier to participation in the cervical screening program. Participation in a cervical screening program has been shown to prevent most cases of cervical cancer (Keppell, Gunn, Hegarty, Madden, O'Connor, Kerse, & Judd, 2003).

Authentic Activities are Ill-Defined, Requiring Students to Define the Tasks and Subtasks Needed to Complete the Activity

In the SET project, we specifically attempted to engage the students by providing cases that the student would need to examine from a clinical

Table 2. Authentic learning design principles and their concrete applicatons within the design of the virtual dental clinic (see Figures 3 and 4)

Principle (Herrington, Oliver, & Reeves, 2002)	**Concrete Application (Virtual Dental Clinic)**
Authentic activities have real-world relevance.	• The clinical case on diabetes was developed in conjunction with experts from medicine and dentistry. • The virtual dental clinic was designed using photographs of the actual clinical setting. Photographs of the clinic were used to build the setting in order to keep the real-world relevance. • Actual case information from real-life patients was used to develop the scenario. This clinical information included actual photographs, radiographs, and medical case information in relation to a child with diabetes.
Authentic activities are ill-defined, requiring students to define the tasks and subtasks needed to complete the activity.	• The diabetes scenario is presented to the students as an actual case. They complete sections on patho-physiology, medical management, and dental management and then apply their knowledge in a realistic case. • There is no step-by-step sequence to completing the clinical case. Students are expected to determine the most appropriate strategies and sequence for completing the treatment plan. • The case information is accessible in any order. The students must make a decision as to the information they require at a particular point in the clinical case.
Authentic activities comprise complex tasks to be investigated by students over a sustained period of time	• The clinical case requires students to concentrate their energy toward the case for a period of approximately 45–60 minutes. • It is expected that the student would return to the case over a period of time as their knowledge is elaborated in the area.
Authentic activities provide the opportunity for students to examine the task from different perspectives, using a variety of resources.	• A wide variety of resources assist the student in obtaining an in-depth examination of the case. • These resources include the following: o Seven clinical photographs o Three radiographs o Patient history o Medical history o Dental history o Height/weight o Social history o Expert opinions from a teacher, psychologist, and endocrinologist.
Authentic activities provide the opportunity to collaborate.	• This module was created as a self-directed learning activity that is used by the professors in a lab setting. • Students complete the module and are provided with expert assistance as required. • Collaboration with other students is not explicit, although teaching staff could complete collaborative group activities at certain points in the tutorial.
Authentic activities provide the opportunity to reflect.	• Explicit reflective activities have not been included in the design of the module. • The use of questions immediately following the presentation of the virtual dental clinic may encourage students to backtrack and re-examine information. • The use of an electronic reflective journal at appropriate points would foster reflection. This feature may be considered at a future time.
Authentic activities can be integrated and applied across different subject areas and lead beyond domain-specific outcomes.	• This case is focused on dental students who will learn generic case-based strategies. • Generic skills of observation, analysis, synthesis, and professional practice would be fostered.
Authentic activities are seamlessly integrated with assessment.	• This is a major strength of the virtual dental clinic. • Students complete a treatment plan that contains all treatment procedures used by an Australian dentist. • The treatment plan encourages the students to re-examine the clinical information (case information, photographs, radiographs) and complete a legitimate treatment plan for the diabetes child. • Students submit their treatment plan and compare their plan to an expert treatment plan. • Students can confirm their treatment plan or re-examine the clinical photographs and radiographs to determine where they may have been incorrect in their initial judgement.
Authentic activities create polished products valuable in their own right rather than as preparation for something else.	• This module is valuable in its own right as a clinical case. Future cases may be developed around the virtual dental clinic. • The clinical case is also the fundamental interaction utilized by the dentist in clinical practice.
Authentic activities allow competing solutions and diversity of outcome.	• Students examine the clinical information and create a treatment plan based on their clinical judgement. • Students may clearly perceive all relevant clinical information and complete the case successfully. However, the design of the case encourages the student to backtrack and re-examine the clinical information and to adjust their treatment plans after submitting to obtain an expert treatment plan.

perspective. Topic 1 examines five mini-cases and provides perspectives on the barriers of cervical screening from the woman's different cultural background. Learners may have insufficient knowledge and will need to examine additional clinical information in the form of a glossary of key medical terms and the library that provides detailed information about cervical screening. This additional information may assist the student in interpreting and analyzing each case more effectively. Students can identify their own knowledge deficiencies and supplement

Table 3. Authentic learning principles and their concrete application in the design of the SET module (see Figures 5, 6, and 7)

Principle (Herrington, Oliver, & Reeves, 2002)	**Concrete Application (Cervical Screening Module)**
Authentic activities have real-world relevance.	• The clinical case on cervical screening was developed in conjunction with experts in medicine from three different universities in Australia and New Zealand. • State health departments involved in cervical screening also participated in the design of the module. • Topic 1 examines five cases from different cultural and ethnic backgrounds and the barriers that differ across these cultures. • Topics 2, 3, and 4 follow one woman into the communication, examination, and follow-up stages. • The virtual clinic was based on an actual clinical setting. Photographs of the clinic were used to build the setting in order to keep the real-world relevance. • Actual case information from real patients was used to build the scenario. This included actual photographs and video and case information in relation to medical history.
Authentic activities are ill-defined, requiring students to define the tasks and subtasks needed to complete the activity.	• The cervical screening cases presented in Topic 1 require the student to examine the different types of barriers of different women from different ethnic and cultural backgrounds. • Students should begin to understand the influence of personal factors, previous experience with cervical screening, and cultural norms in relation to treating different women in the clinical setting. • Questions are open-ended, which requires the student to examine a variety of information resources to determine appropriate responses. • The case information is available for students to access as needed. A glossary and library provide additional resources for examination.
Authentic activities comprise complex tasks to be investigated by students over a sustained period of time	• The clinical case requires students to concentrate their energy toward the case for a period of approximately 3–4 hours. • It is expected that the student would return to the case over a period of time as their knowledge is elaborated in the area. • The video examples of communication and examination provide a means for the student to view best-practice examples before clinical practice and as revision after clinical practice. Used in these two ways, the student should be able to reflect on appropriate strategies for improving the clinical practice of this sensitive area.
Authentic activities provide the opportunity for students to examine the task from different perspectives, using a variety of resources.	• A wide variety of resources assist the students in obtaining an in-depth examination of the case. • These resources include the following: o Clinical photographs o Video of ideal communication between the doctor and the woman o Video of ideal examination procedures o Patient history o Medical history o Social history o Expert opinions from practicing doctors o Definitions of key words in the glossary o Extended information in the library o Expert feedback
Authentic activities provide the opportunity to collaborate.	• This module was created as a self-directed learning activity and as an adjunct to the cervical screening program in Victoria. • It is expected that participants in the clinical screening program will discuss key issues and concerns using aspects of the module as a trigger for activities. • Collaboration with other students is not explicit at this point in time, although teaching staff could complete collaborative group activities at certain points in the tutorial.
Authentic activities provide the opportunity to reflect.	• Explicit reflective activities have been included in the design of the module. • A reflective notebook allows the students to document reflections and ideas throughout the module. This can be saved as an electronic file or printed out at the end of the session. The use of an electronic reflective journal at appropriate points would foster reflection. • Open-ended questions require the student to complete a detailed response before proceeding. This can be saved as an electronic file for future examination. • The use of questions immediately following the presentation of the video segments may encourage students to backtrack and re-examine information.
Authentic activities can be integrated and applied across different subject areas and lead beyond domain-specific outcomes.	• The module examines a difficult and sensitive area in which students may experience some awkwardness. • It is hoped that students will begin to learn skills in empathy and improve communication skills that can be applied to other sensitive areas in medical clinical practice, such as breast examinations. • Generic skills of observation, analysis, synthesis, and professional behavior would be fostered.
Authentic activities are seamlessly integrated with assessment.	• Formal assessment has not been determined. Future implementation will determine this assessment. • Self-assessment activities are embedded throughout the module. Students are asked to complete activities after viewing each video segment.
Authentic activities create polished products valuable in their own right rather than as preparation for something else.	• This module is valuable in its own right as a clinical case. • The clinical case is also the fundamental interaction utilized by the medical doctor in clinical practice.
Authentic activities allow competing solutions and diversity of outcome.	• Students examine the clinical information and create a schema for future medical consultation. • The design of the case encourages the student to backtrack and re-examine the clinical information.

their ideas via the additional resources provided (see Figure 8).

Authentic Activities Comprise Complex Tasks to be Investigated by Students over a Sustained Period of Time

SET will be used by universities in Australia and New Zealand to introduce the concepts of cervical screening and intimate examinations in a consistent way that emphasizes the importance of sensitive communication and examination skills. It should be noted that many students are nervous about performing intimate examinations, and the SET module will attempt to demystify the clinical process. Because the module focuses on communication skills in conjunction with clinical skills, students will need to examine the module over a period of time. Initially, students may obtain an overview of the cervical screening pathway before undertaking clinical practice. They may also revise certain sections after clinical practice. While practicing as a general practitioner, the doctor may revisit appropriate sections of the module and examine best-practice processes

Figure 3. Virtual dental clinic

Figure 4. Virutal dental office

Figure 5. Situated learning environment utilized for the SET Project–1

Figure 6. Situated learning environment utilized for the SET Project–2

and procedures in the field. For this reason, the four to five hour module will require sustained attention over a period of time and for different purposes. Specifically, the SET module will be utilized in different ways in the medical program of Melbourne, Queensland, and Auckland. The module will be used to compliment face-to-face teaching as well as allow students to explore the "learning loops" provided on SET both in the tutorial and self-directed learning (SDL) setting. The SET module will be available for students to use in the computer labs/PBL rooms at each site.

Authentic Activities Provide the Opportunity for Students to Examine the Task from Different Perspectives, Using a Variety of Resources

The use of different cases provides the students with an ability to examine different barriers

Figure 7. Situated learning environment utilized for the SET Project–3

Figure 8. Glossary utilized in the SET module

that may affect cervical screening. The library resources also provide additional information and a variety of resources that should assist the student (Herrington, Oliver, & Reeves, 2002) to "examine the problem from a variety of theoretical and practical perspectives, rather than allowing a single perspective that learners must imitate to be successful" (p. 281) (see Figures 9 and 10).

Authentic Activities are Ill-Defined, Requiring Students to Define the Tasks and Subtasks Needed to Complete the Activity

Both modules examine complex content areas in the areas of dentistry and medicine. Due to the complexity of the content, a range of resources pro-

vides definitions of key terms and more elaborate content in the library. Key dental, medical, and physiological concepts are defined in a glossary. Users can choose to browse all glossary terms or select specific terms that require clarification at the relevant point in the module. In addition, resources are provided in a library section in the SET module. The virtual dental clinic has a range of clinical images and case information available within the clinical setting, allowing students to access as required in the examination of the dental case. We specifically attempt to engage the students by providing cases that the student will need to examine from a clinical perspective. This means that they may need to search for relevant content in other sections of the module to achieve a deep understanding of the content.

Authentic Activities Comprise Complex Tasks to be Investigated by Students over a Sustained Period of Time

Each module focuses on 1 to 3 hours of difficult dental, clinical, and physiological content that is used to complement face-to-face teaching. The modules will also be used to support lectures by allowing lecturers to recommend that students examine the self-directed learning module located in a lab setting. The students may also be asked to utilize the modules as a revision tool before examinations. Because the modules focus on conceptually difficult content, they will require multiple exposure and viewing by the students. For example, students can review communication practices before they complete cervical screening in a clinical setting and then after clinical practice to review their communication and examination methods. Dental students can reinforce clinical treatment protocols in treating a pediatric dental patient by working through a legitimate case on diabetes.

Authentic Activities Provide the Opportunity for Students to Examine the Task from Different Perspectives, Using a Variety of Resources

Within the virtual dental clinic, we provide multiple entry points into the clinical information. This allows the user to explore the clinic and obtain the necessary clinical information about the patient. We wanted the student to "criss-cross the landscape of knowledge" in order to obtain information that will enable them to complete the relevant treatment plan (Young, 1993, p. 46). The user can navigate around the clinic and find information in two ways. By traveling around the clinic, they will find hotspots that provide clinical information. A site map also provides clinical information enabling the student to access patient records (patient information, medical history, height and weight, dental history, and social history), seven clinical slides, three radiographs, and expert information from a teacher, psychologist, and endocrinologist. By providing the necessary scaffolding for the novice learner, we attempt to move the learner toward expert case management (see Figure 11–14).

The use of different cervical screening cases provides the students with an opportunity to examine different barriers that may affect cervical screening. The library resources also provide additional information and a variety of resources that should assist the student (Herrington, Oliver, & Reeves, 2002) to "examine the problem from a variety of theoretical and practical perspectives, rather than allowing a single perspective that learners must imitate to be successful" (p. 281).

Authentic Activities Provide the Opportunity to Collaborate

Within the virtual dental clinic, the learning context allows the student to learn within an expert-supported learning environment. Lab

Figure 9. Clinical visuals and radiographs that can be enlarged for further detail within the virtual the virtual dental clinic

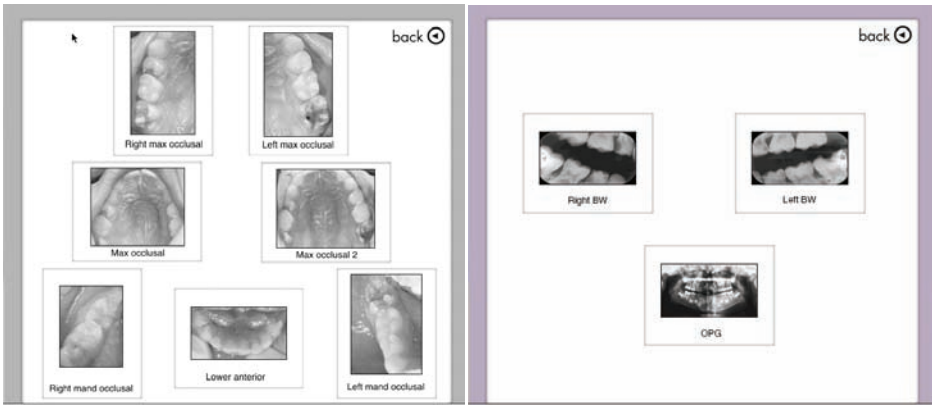
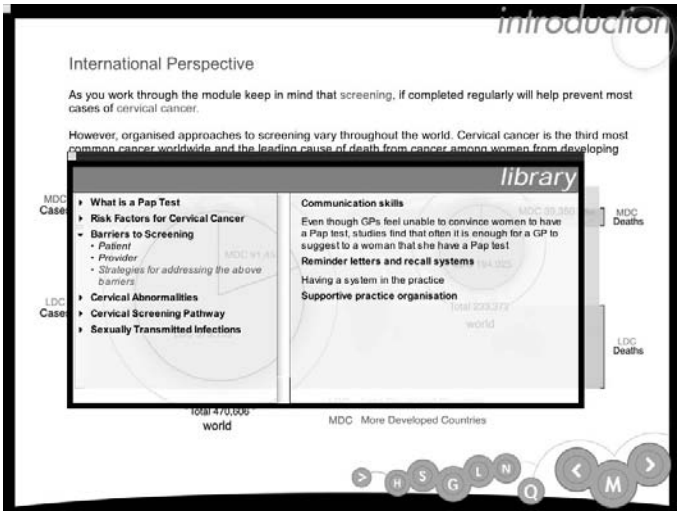


Figure 10. Library resources utilized by the student for additional clinical information about cervical screening



sessions have been scheduled in which students interact with the module and were supported by pediatric dentists. According to Young (1993), "From the perspective of situated cognition, the teacher's role should be to 'tune the attention' of students to the important aspects of the situation or problem-solving activity, specifically those attributes that are invariant across a range of similar problems and therefore will transfer to many novel situations" (p. 47). The advantage of utilizing expert support is that some information about the case can be clarified and explained by the expert tutors. Their roles in the learning process are as a coach, collaborator, and mentor for student learning. According to Choi and Hannafin (1995), coaching focuses on "directing

Figure 11. Student response to an open-ended question with an expert answer within the SET project

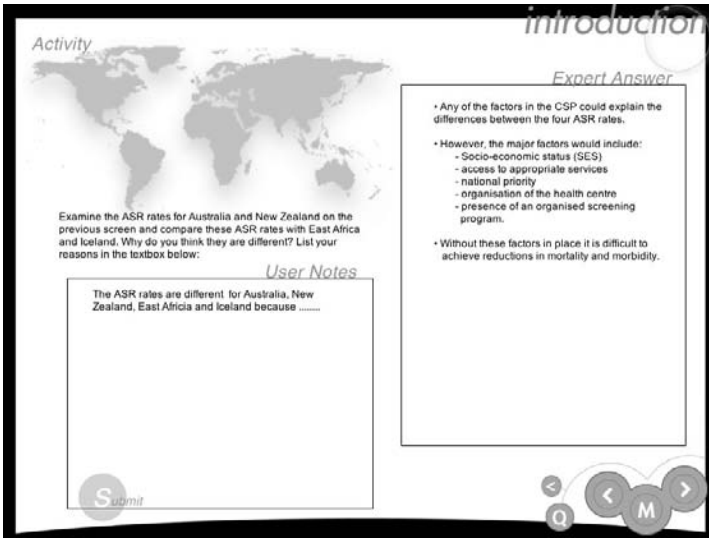
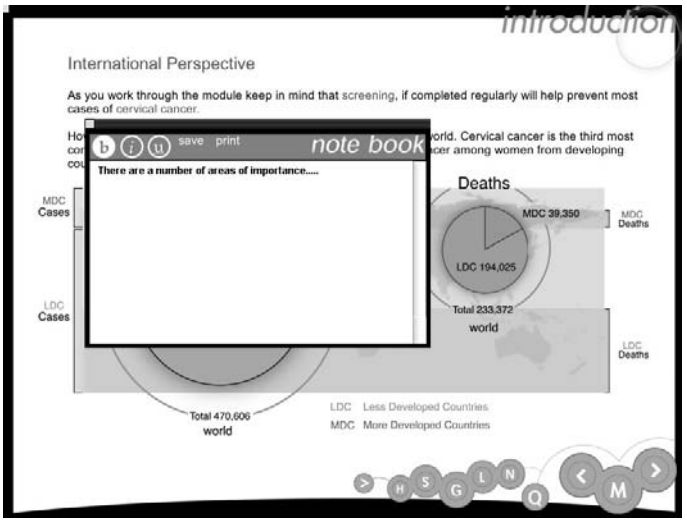


Figure 12. Reflective notebook utilized in the SET project



learner attention, reminding of overlooked steps, providing hints and feedback, challenging and structuring ways to do things, and providing additional tasks, problems, or problematic situations" (p. 62). The SET module will also be used in a collaborative setting.

Authentic Activities Provide the Opportunity to Reflect

Many of the activities embedded throughout the SET module ask the student to type in extended responses to questions. These responses are then submitted, and the student is provided with expert

Figure 13. Electronic treatment plan utilized for the diabetes clinical case

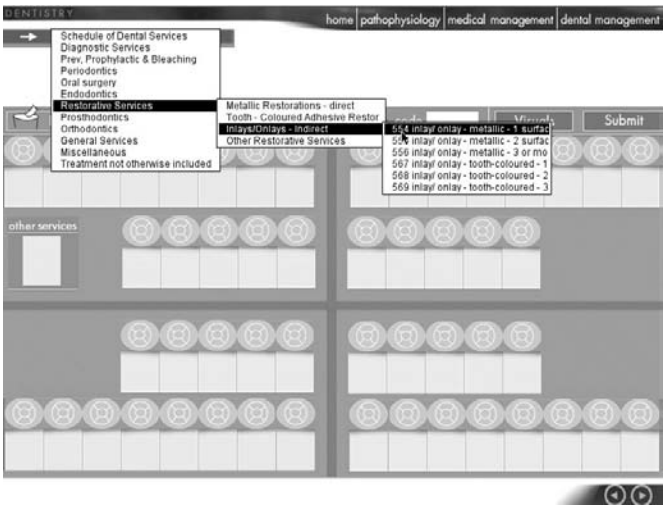
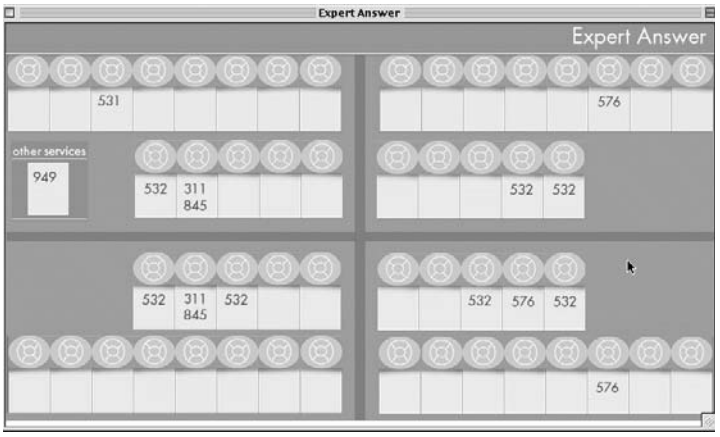


Figure 14. Expert answers for the diabetes treatment plan



feedback. In addition, these typed responses can be saved by the student and referred to at a later point in their study. This approach was adopted to encourage students to engage with the content and synthesize ideas in their own words. Students are also encouraged to write their own notes and ideas in their personal notebook that can be printed or saved as a Microsoft Word file for review at a later point in the semester.

Authentic Activities can be Integrated and Applied Across Different Subject Areas and Lead Beyond Domain-Specific Outcomes

Students will be able to view the cervical screening cases from different perspectives. Medical students, nurses, and other health professionals should become aware of different disciplinary approaches and roles in cervical screening. Students

should begin to see the cervical screening process from the woman's perspective as opposed to the clinical perspective.

The case-based approach utilized in the virtual dental clinic allows students to examine a process for examining other similar cases in dentistry.

Authentic Activities are Seamlessly Integrated with Assessment

Young (1993) suggested that traditional forms of assessment in situated learning may prove to be inadequate. Authentic learning tasks must be assessed using methods that best align with the task. Within the virtual dental clinic, we attempted to do so by asking the students to develop a treatment plan using an electronic version of the protocols utilized in a traditional clinic. Our aim in developing this method of assessment was to foster higher-order thinking skills. We were also conscious of providing a method of formative assessment that allowed (Choi & Hannafin, 1995), "generation of ideas and the presentation of problem-solving processes such as planning, implementing, and revising" (p. 65). In order to simulate the use of the treatment chart, an electronic dental chart was created to allow users to create treatment plans. The electronic dental chart allows the trainee dentist to select a category of dental services and allocate a specific treatment for individual teeth. This chart also allows the user to complete cases and submit information to obtain expert feedback. The expert feedback provides precise information about the treatment protocol for each tooth. The students can compare their treatment plans to the experts' plans and then reenter the virtual clinic to rectify any inaccuracies and misconceptions. This concept of authentic assessment tasks is essential in the instructional design of authentic learning tasks. As stated in Young (1993), "Assessment should be a seamless, continuous part of the activity (a learning/assessment situation)" (p. 48). The final exam paper also mirrors the authentic learning task in the multimedia package by utilizing a case-based learning approach.

Although not a formal part of assessment, the SET module will be gradually utilized in all parts of the medical curriculum. Activities have been embedded that reflect real-world decisions and clinical encounters. Students will be given the opportunity to complete a number of on-screen sequencing and clinical interpretation tasks. To assess whether they are able to identify the barriers to cervical screening, they will be required to respond to a number of questions relating to the video sequence. Student assessment will be via the above self-assessment activities and tutor feedback during the tutorial setting. As the assessment is developed, we will utilize principles as outlined by Young (1993).

Authentic Activities Create Polished Products Valuable in Their Own Right rather than as Preparation for Something Else

Both modules are polished products in their own right, which focus on specific content relevant for treatment in dentistry and cervical screening. By studying each of the modules, the students will learn in-depth knowledge about the topic. Because the modules focus on a difficult content area, they act as stand-alone modules or complement other aspects of the curriculum. For instance, the SET module will be utilized to complement the existing forms of teaching cervical screening, which include face-to-face settings, video, pelvic models, observation, and clinical practice.

Authentic Activities allow Competing Solutions and Diversity of Outcome

The virtual dental clinic focuses on the examination of clinical information to develop a treatment protocol, whereas the SET module emphasizes communication, process, and procedures in relation to cervical screening. In examining the SET

module, students should be able to learn about communication practices that can be transferred to other clinical areas, particularly, of a medically sensitive nature. Students should be able to interpret content in a diversity of ways and begin to assimilate some of these concepts into their own process of completing cervical screening.

EVALUATION

An initial evaluation was undertaken to determine user perceptions in relation to the virtual dental clinic. Three practicing pediatric dentists participated in a focus group. A number of insights were gained into the design of the virtual clinic and its authenticity. We examined the match between the actual clinical setting and the virtual dental clinic. It appeared that presenting all relevant clinical information in one session may overwhelm the student:

If we think about the diabetes case and ... the virtual clinic, what do you think about that case in terms of how realistic it was…

It would be quite different. You would probably have more than one appointment and that was quite tricky deciding—on the first day I probably wouldn't do all these things. You could put them under general anaesthetic. Apart from that it was quite difficult if you had to write the whole lot out on the same day you would probably do something after the general anaesthetic get them back two weeks later and say okay, how's the toothbrush going.

Navigation also appeared to be an issue in the virtual clinic. We provided two methods of navigation. These included a site map and hotspots in the clinical setting. Although we provided a brief tutorial for the students, all three students still failed to utilize the hotspot information.

When I moved the mouse it came up and I didn't realise what I could click. I think if you want us to get the picture on the top of the menu, you should have it flashing or something like that. So that it shows up.

Further design work is being undertaken to address this misunderstanding. An animation will be utilized instead of the existing help screens to highlight relevant information in the virtual clinic. Our design attempted to provide an open exploration of the clinic. However, this proved too advanced for the users. Although we provided flexible access to the resources to solve the clinical case, the students still utilized the site map in a traditional top-to-bottom and left-to-right reading pattern as opposed to clicking on information that they considered relevant at that point in the case:

When you went into the office?
I went back to the menu and went to help. I couldn't understand. So then I went back to the other picture and found it by accident. I clicked on the telephone or something.

The sequence of information was also important to the user:

Did you actually read about the teacher; the specialist information?
Don't you think this information should go before the examination? You are talking about the medical history and social history. It should go before the examination.
If it was a referral, you would get a note from a doctor before she actually comes in, which gives you the whole..., and you don't have to look for anything.

Our goal of designing the module so that students would need to revisit some clinical information appeared to be successful in some instances:

Did you go back to the visuals when you were completing the treatment plan?
Did you need any other information at that point?
Patient file, complaints and history. I did use it, but I did go back to the pictures quite a lot. How many times? Three, five times. I think I wrote it down which made it a bit easier. I don't know the numbers so it made it a lot harder.

However, it appeared that the students had difficulty providing a treatment number from the electronic treatment plan as opposed to the usual paper-based booklet. It also appeared that obtaining expert feedback could have been optimized:

Did you all submit and get the expert answers? I didn't get the answers. I went to next. That's why I didn't find the expert…. Maybe it's a good thing to put the expert under next. Down the bottom. So you go to the next phase.

Further evaluation with a subsequent group should also inform the design and provide feedback to the development team. However, it appears that there is a fine line between authentic myth and reality.

Although the SET module has not been formally evaluated, our next step in the process is to complete extensive evaluation on the module from different perspectives. In the first instance, we will evaluate the module with clinicians experienced in the process. Some formative evaluation has already been undertaken with this group. In the second instance, a number of multimedia experts will be asked to evaluate the learning design of the module. Student evaluation will be undertaken over the next 12 months. We are also examining the evaluation of the module in cross-cultural settings to determine its applicability for different international usage. The module examined one woman's journey through the cervical screening process. In the future, we intend to examine modules for different ethnic groups.

Major Problems Encountered

The examination of the two modules within this chapter demonstrates a student-centered approach to their design. They represent rich teaching and learning resources that address conceptually difficult content areas. However, there are also cases of projects that have not been successful in developing viable educational resources. We cannot assume that design and development processes adopted by many development teams are sufficient to assure the success of a project. Burford and Cooper (2000) support this view and suggested that "most academics are not skilled in interface design, multimedia, selecting appropriate technologies for online teaching, nor project management" (p. 207). For this reason, they suggested that "whatever the developmental model, it must be achievable within the confines of established practice and available resources" (p. 209). However, it is important to remember that online and multimedia design and development are highly creative processes. Design and development models help to minimize potential difficulties in projects, but they never eliminate all constraining factors. According to Burford and Cooper (2000), "Whatever the model used, however well defined the process, the development process itself is a diffuse and difficult one. From its first conception to a final product, a development process is fraught with undefinable influences and unpredicted contributing factors" (p. 210). This section examines some of the consistent problems encountered in the transformation of a curriculum with multimedia and online learning:

- A consistent problem encountered in designing and developing multimedia modules is that academic staff members are over ambitious with their goals. Through a process of coaching and educating academic staff, it is possible to change this perception. Because it often requires 300–800 hours of development per course hour, multimedia should only be used for areas where it is

appropriate. The virtual dental clinic and the cervical screening module represent conceptually difficult content that warrant the above time and effort on the part of the development team. The first two questions we ask are as follows: Can this be completed using another teaching method? And why multimedia? Technology must enhance the teaching and learning process and should be used for addressing learning misconceptions and complex and difficult content that cannot be easily explained in another form.

- Another common problem is that many academic staff members apply their traditional teaching styles in designing and developing multimedia and online learning. Academic staff often focus on instructivist teaching models in developing their online or multimedia modules. The challenge in this situation is for the instructional designer to provide other pedagogical perspectives and suggest alternative approaches such as case-based reasoning, PBL, and authentic learning environments. There are a number of factors that need to be considered in this situation. For instance: What are the learning outcomes of the module? What pedagogical methods best suit the attainment of these outcomes? It is important to always take an eclectic approach to the teaching and learning process and suggest methods of teaching best suited to the entire learning context. There is a delicate balance between implementing the content expert's approach and coaching the academic in other pedagogical possibilities, which may enhance the learning of the content by the user.
- It is important not to underestimate the time required to design and develop online and multimedia products. As suggested by the above formulas, there must be an excellent pedagogical reason for justifying the resources required for multimedia and online learning.
- Working in teams is both a rewarding and challenging experience. It is important to employ a dedicated manager who can coach the design and development team in appropriate processes, and who has the ability to harness creative energy and deal with the inevitable tensions that arise in creative teams. By focusing on the goal as opposed to personal ownership, tensions can often be dissipated.

A potential bottleneck exists between the content expert and design and development staff (instructional designers, graphic designers, programmers, etc.) in terms of translating content into a form that embodies sound educational design. A process or strategy is required to streamline the interaction between the instructional designer and subject matter expert (Keppell, 2001). One of the most important principles in any project is to clarify the roles and expectations of the client/SME. Many projects fail due to an inappropriate consideration of what the client/SME expects from the project. According to Coscarelli and Stonewater (1979–1980), "An understanding of client psychological types and an ability to differentially respond to various types is a particularly effective designer strategy for relationship building and managing" (p. 16). It is therefore essential to establish a successful working relationship with an SME by determining philosophical assumptions of the SME before beginning the instructional design (Davies, 1975), as "a great deal of what is accomplished depends on the quality of the client–consultant relationship" (p. 351).

CONCLUSION

Learner-centered approaches emphasize problem-based and case-based learning, which attempt to create realistic learning interactions in order to replicate professional practice. The above discussion demonstrates how authentic

learning multimedia modules have been used to complement a medical, dental, and health science curricula. The learning design for each module is carefully articulated in order to provide an insight into the process of instructional design that can be transferred to other learning settings and other content areas. The articulation of the learning design also demonstrates the explicit instructional design decisions that were made in the two modules, each of which are based on a constructivist teaching and learning model. It is also suggested that interactive, media-rich multimedia modules are relevant for difficult content areas or areas where misconceptions may be prevalent in the curriculum. The articulation of this model of teaching and learning allows other designers and researchers to examine the applicability of these designs for their own setting and circumstances.

ACKNOWLEDGMENT

The author wishes to acknowledge the graphic designers on both projects: Jennifer Kirk, Avril Martinelli, Jacqui Jewell, Andrew Bonollo, and Carolyn Casey from the Biomedical Multimedia Unit.

REFERENCES

Altschuld, J. W., & Witkin, B. R. (2000). *From needs assessment to action: Transforming needs into solution strategies.* Thousand Oaks, CA: Sage Publications, Inc.

Burford, S., & Cooper, L. (2000). Online development using WebCT: A faculty managed process for quality. *Australian Journal of Educational Technology, 16*(3), 201–214.

Choi, J. -I., & Hannifin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology, Research and Development, 43*(2), 53–69.

Coscarelli, W. C., & Stonewater, J. K. (1979–1980). Understanding psychological styles in instructional development consultation. *Journal of Instructional Development, 3*, 16–22.

Davies, I. K. (1975). Some aspects of a theory of advice: The management of an instructional developer–client, evaluator–client, relationship. *Instructional Science, 3*, 351–373.

Herrington, J., Oliver, R., & Reeves, T. C. (2002). Patterns of engagement in authentic online learning environments. In A. Williamson, C. Gunn, A. Young, & T. Clear (Eds.), *Winds of change in a sea of learning. Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Tertiary Education* (pp. 279–286). Auckland, New Zealand: UNITEC: Institute of Technology.

Jonassen, D. (1996). *Computers in the classroom: Mindtools for Critical Thinking.* Englewood Cliffs, NJ: Merrill.

Jonassen, D., & Land, S. M. (2000). Theoretical foundations of learning environments. Mahwah, NJ: Lawrence Erlbaum.

Jonassen, D., & Reeves, T. (1996). Learning with technology: Using computers as cognitive tools. In D. Jonassen (Ed.), *Handbook of research on educational communication and technology* (pp. 693–719). New York: Scholastic.

Jonassen, D., Mayes, T., & McAleese, R. (1993). A manifesto for a constructivist approach to uses of technology in higher education. In T. M. Duffy, J. Lowyck, & D. H. Jonassen (Eds.), *Designing environments for constructive learning.* Berlin: Springer-Verlag.

Keppell, M. (2001). Optimising instructional designer—Subject matter expert communication in the design and development of multimedia proj-

ects. *Journal of Interactive Learning Research, 12*(2/3), 205–223.

Keppell, M., Gunn, J., Hegarty, K., Madden, V., O'Connor, V., Kerse, N., & Judd, T. (2003). Using authentic patient interactions to teach cervical screening to medical students. In D. Lassner, & C. McNaught (Eds.), *Proceedings of ED-Media 2003 World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 1431–1438). Honolulu, Hawaii, Association for the Advancement of Computing in Education.

Keppell, M., Kan, K., Brearley Messer, L., & Bione, H. (2002). Authentic learning interactions: Myth or reality? In A. Williamson, C. Gunn, A. Young, & T. Clear (Eds.), *Winds of change in a sea of learning. Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Tertiary Education* (pp. 349–358). Auckland, New Zealand: UNITEC: Institute of Technology.

Keppell, M., Kennedy, G., Elliott, K., & Harris, P. (2001, April). Transforming traditional curricula: Enhancing medical education through multimedia and web-based resources. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning (IMEJ), 3*(1). Retrieved March 2002 from the World Wide Web: http://imej.wfu.edu/articles/2001/1/index.asp

Koschmann, T., Kelson, A. C., Feltovich, P. J., & Barrows. H. S. (1996). Computer-supported problem-based learning: A principled approach to the use of computers in collaborative learning. In T. Koschmann (Ed.), *Computer supported collaborative learning: Theory and practice in an emerging paradigm*. Mahwah, NJ: Lawrence Erlbaum.

Kozma, R. B. (1987). The implications of cognitive psychology for computer-based learning tools. *Educational Technology, 27*(11), 20–25.

Suivinen, T., Messer, L. B., & Franco, E. (1998). Clinical simulation in teaching pre-clinical dentistry. *European Journal of Dental Education*, (2), 25–32.

Young, M. (1993). Instructional design for situated learning. *Educational Technology Research and Development, 41,* 43–58.

This work was previously published in Interactive Multimedia in Education and Training, edited byS. Mishra and R.C. Sharma, pp. 350-376, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.6
On a Design of SCORM-Compliant SMIL-Enabled Multimedia Streaming E-Learning System

Sheng-Tun Li
National Cheng Kung University, Taiwan, ROC

Chu-Hung Lin
National Sun Yat-sen University, Taiwan, ROC

Pao-Ta Yu
National Chung Cheng University, Taiwan, ROC

ABSTRACT

E-learning plays a key enabling role in knowledge management for individuals, schools, and enterprises. Nevertheless, the lack of standards in content and learning management systems (LMSs) makes the reusability and interoperability of learning resources infeasible. The emergence of the SCORM specification has shed light on the standardization of e-learning. Unfortunately, the existing SCORM-compliant asset model is simplified; only a few asset types are allowed. On the other hand, W3C's declarative-style SMIL (Synchronized Multimedia Integration Language) is becoming prevalent in designing Web-based instructions with the consideration of temporality and spatiality of presentations. In addition, advances in real-time multimedia technologies can vitalize further these instructions. In this study, we propose an SMIL-enabled asset model with the enhancement of multimedia streaming. To render SCORM-compliant SMIL-enabled streaming contents, a Java applet-based SMIL RTP/RTSP LMS system is designed. We develop the proposed system in the Java EJB environment to tackle the issue of platform interoperability. The

resulting system demonstrates an encouraging direction towards a more vivid and interactive SCORM-compliant e-learning.

INTRODUCTION

The rapid expansion of the Internet has had a dramatic impact on both our lives and our education. The major feature that differentiates e-learning from traditional learning is its ability to train anyone, anytime, and anywhere, which is attributed to the openness of the Internet. Without temporal and spatial limitations, a person can have an independent and individual learning space. Universities first realized the advantages of e-learning and actively built up the so-called cyber universities that offered diverse asynchronous or synchronous distance learning in regular, professional, and/or continued educations. Next, business companies also recognized the significant market benefits of e-learning. Analysts predicted that corporate spending on e-learning programs would reach US $23 billion by 2004. Yet, some innovative enterprises have moved beyond training to focus on e-learning in the context of knowledge management and found that e-learning is a vital step in the development of knowledge management systems (Ismail, 2002).

The major feature of e-learning is digitalized courseware, which is managed by the so-called Learning Management System (LMS). In addition, LMS helps learners to learn the coursewares and keep track of their learning behaviors. But coursewares in different LMS platforms cannot be interoperable directly, so the attempt to share learning resources is impeded. The most critical issue for this is whether the process of designing the coursewares follows some kind of standards. Besides, the heterogeneity existing in the different LMSs is another impediment. There is no way to monitor and evaluate the learner's behavior from one platform to another. All these hinder the sharing of learning resources, including platforms and contents. Although it has been recognized that developing reusable and sharable content objects is of great importance (Anido et al., 2001a, 2001b; Muzio et al., 2002; Rifon et al., 2001), not all of the aforementioned issues have been addressed. To overcome these hurdles, a variety of e-learning standards have been developed, such as IMS, IEEE LTSC, AICC CMI, LMML, ARIADNE, and ULF, and so forth. Recently, Advanced Distributed Learning (ADL) organization, supported by the U.S. Department of Defense, and major e-learning vendors, established one specification called Sharable Content Object Reference Model (SCORM) toward standardizing e-learning (SCORM, 2003). SCORM integrates and refers to the aforesaid standards from AICC, IMS, IEEE, and several e-learning factories. The objective of the specification is to facilitate the interoperability between SCORM-compliant contents and SCORM-compliant LMS, and to make the valuable resources become durable, interoperable, accessible, and reusable.

On the other hand, recent progress in multimedia technologies is greatly changing people's life styles. The marriage of the Web and multimedia technologies results in the Web-based multimedia presentations, so people can read multimedia lectures via general browsers. However, designing multimedia presentations is not a trivial task. Synchronized Multimedia Integration Language (SMIL) (SMIL, 2004) built on XML and proposed by W3C, provides a simple way to design multimedia presentations in a manner similar to HTML documents. The SMIL specification meets three requirements of multimedia document models: temporal, spatial, and interaction.

Furthermore, the demand for real-time and robust multimedia applications is rising dramatically, as Internet connection speeds improve. Various protocols for supporting networked multimedia services have been proposed. Two of the representatives are Real-Time Transport Protocol (RTP) and Real-Time Streaming Protocol (RTSP). RTP is designed specifically to transmit real-time

media data and enables client applications to play media over the Internet without having to download the entire file at once. RTSP, a state-oriented protocol, provides VCR-like functionality to allow more interaction between users and media sources.

Multimedia presentations can attract learners' attentions because of the vivid video images and sound effects. According to the SCORM specification, multimedia materials, such as images, are defined as *assets* in the content. Currently, however, the samples in this specification or mostly SCORM-complaint LMSs that only handle simple assets like pictures. To vitalize the asset contents of SCORM, this study designs a new asset model based on SMIL and RTP/RTSP standards, so that the features of spatiality, temporality, interaction, and streaming control can be added into assets. The extended asset model conforms to SCORM, and, thus, any SMIL-enabled asset can be imported to LMS, and communicates with LMS to track the learners' learning progresses.

In addition to facilitating the interoperability of sharable course contents, there is a great necessity for making LMS interoperable. The prototype of ADL sample LMS 1.2 environment is oversimplified and cannot fit the requirement of interoperability and reusability. On the other hand, component-based computing has been recognized as a successful paradigm that allows two or more software components to cooperate in a seamless manner, despite heterogeneities in implementation of languages, service interfaces, and deployment platforms (Li, 2002). Therefore, we choose the emerging Enterprise Java Bean (EJB), a distributed component-based computing environment, to develop the proposed LMS for realizing SCORM-compliant SMIL-enabled multimedia streaming contents.

This paper is organized as follows. First, it gives a brief review about e-learning standards. Then, we propose the SMIL-enabled asset model with multimedia-streaming functionality and show how SMIL documents can be wrapped as SCORM assets in the manifest and the structure of SMIL documents. The next section discusses the rendering of SMIL-enabled assets by a Java applet-based SMIL player. Then comes the design of SCORM-compliant LMS in the EJB environment, and after that, a content sample and its presentation on a Web browser are demonstrated. Finally, we conclude this paper by providing some directions about future work.

STANDARDS IN E-LEARNING

In this section, we briefly review several popular standards in e-learning.

Existing Standards

Instructional Management System (IMS) (AICC, 2003; IMS, 2003) is a global learning consortium to develop an open specification for sharing of heterogeneous platforms, reusability of valuable contents, tracking of learned material, exchange of records, and so forth. IMS is the generator of Learning Object Metadata (LOM), which is extended continually to IEEE LTSC. IMS submits suggestions and maps the metadata concept in the so-called content aggregation model (CAM) to IEEE LTSC. Metadata provides a shortcut for content developers to search and retrieve desired contents in the content repository. IMS defines nine basic categories of metadata: general, life-cycle, meta-metadata, technical, educational, rights, relation, annotation, and classification. In addition, IMS introduces the package type of instruction: content packaging. By assembling all physical files and describing the organizations and relations of the content items in a manifest file, a content package is flexible to be assembled, disassembled, merged, and recreated from contents. However, IMS tends to concentrate on the conceptual layers, and lacks the details of implementation.

Learning Technology Standards Committee (IEEE LTSC) (LTSC, 2004) has devoted itself to the tutorial standards of software, tools, development, maintainability, and contents. There are many teams in LTSC, of which the LOM team is responsible for developing the contents' standards. The main target of LOM is to facilitate the sharing and exchange of contents by defining plenteous metadata in LOM.

Computer Management Instruction (CMI) in Aviation Industry Computer-Based Training Committee (AICC) (AICC, 2003) defines the way to construct a course. In addition, CMI provides a data model to define the mechanism of communications in LMS, the status of instruction resources, and the tracking of learners' behaviors. However, from the viewpoint of the data model, CMI lacks the course structure, metadata, and content packaging.

Learning Material Markup Language (LMML) (LMML, 2004) was developed by Passau University in Germany. LMML, an XML-based language, is used in describing course contents; however, it is difficult to add a low-level media and, thus, results in the lack of clear hierarchy.

Alliance of Remote Instructional Authoring and Distributed Networks for Europe (ARIADNE) (ARIADNE, 2004) is the alliance of many universities in different European countries. ARIADNE emphasizes the integration in multi-languages and multi-culture. It generally follows the LOM specification but with minor simplification.

Universal Learning Format (ULF) (ULF, 2002), which originated from XML and RDF, defines content types from a series of learning systems by SABA Inc. It provides a complete specification of description and bridges other standards, such as IMS, ADL, IEEE LTSC, and so forth.

SCORM

Sharable Content Object Reference Model (SCORM) adopts the concepts of metadata and content packaging from IMS and IEEE LTSC to define a standard set of metadata element definitions that can be used to describe learning resources. SCORM Version 1.2, announced by ADL, is composed mainly of Content Aggregation Model (CAM) and Run-Time Environment (RTE). CAM focuses on how to construct reusable and sharable contents, whereas RTE ensures the interoperability of different LMSs.

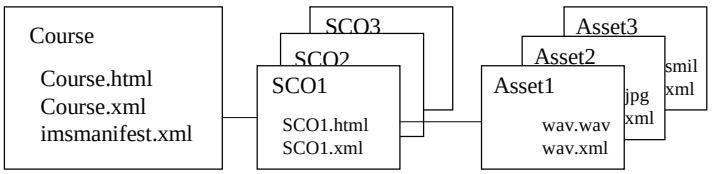
Content packaging defines the structure and the intended behavior of a collection of learning resources. It is used to provide a standardized way to exchange digital learning resources among different systems or tools. A package uses a manifest file that contains the following information for describing the package:

- meta-data about the package
- an optional organization section that defines content structure and behavior
- a list of references to the resources in the package

In a content package, the organizations component defines the structure of the content, and the resources component describes external resources as well as the physical files of which the package consists. The resources are referred to at various points within the organization component. The type of the SCORM resource is defined by the attribute, named "adlcp:scormtype". The value of the "adlcp:scormtype" is either "asset" or "sco".

An asset is an electronic representation of media, text, image, sound, Web page, assessment objects, or other pieces of data deliverable to Web clients. Learning contents normally are composed of assets. On the other hand, a Sharable Content Object (SCO) contains a collection of one or more assets that include a specific launchable asset utilizing a SCORM RTE to communicate with LMS. A SCO represents the lowest level of granularity of learning resources that can be tracked by an LMS in a SCORM RTE. Figure 1

Figure 1. Concept diagram of SCORM course content



shows the concept diagram of a SCORM course content. Each course is composed of several SCOs, which, in turn, are constituents of several assets. The structure of the course content is defined in imsmanifest.xml_whereas the metadata of each learning resources is described by the corresponding XML files.

The purpose of SCORM RTE is to provide a means for interoperability between SCO-based learning contents and the learning management systems. An essential requirement of SCORM is that learning contents should be interoperable across multiple LMSs, regardless of the tools used to create them. For this to be possible, there must be a common way to start the content and a common way for the content to communicate with an LMS and predefined data elements that are exchanged between an LMS and the content during its execution. The three components of the RTE are defined in SCORM as Launch, Application Program Interface (API), and Data Model.

SMIL

SMIL provides a simple way to design TV-like multimedia presentations on the Web (SMIL, 2004). The SMIL syntax conforms to the XML standard, so one can write SMIL presentations using general text editors. There are several salient features that make SMIL attractive in designing Web-based multimedia presentations.

- The presentation designer can indicate the spatial layout and temporal relationships of media objects.
- Media objects can be distributed either from the same resource of the SMIL document or any place specified by the URL.
- It supports the functionality of hyperlinks.
- Media objects can be presented on a client's browser according to system and personal preferences, such as network bandwidth, user language, and screen resolution.

An SMIL presentation usually is rendered by a specific player, such as RealPlayer. Unfortunately, most Web browsers, including Microsoft Internet Explorer 6.0 and Netscape, do not support SMIL documents yet. The rendering work involves pre-fetching and buffering media objects. To accommodate the requirement of SCORM, we extend our previous work on SMIL players (Li & Chen, 2001) and propose a Java applet-based architecture for rendering SMIL-enabled assets on IE in this study.

Multimedia Streaming

RTP, defined in IETF RFC 1889, is a transport protocol for delivering real-time multimedia streams over multicast or unicast IP networks. RTP usually is carried by UDP in order to efficiently transmit data under time constraints; thus, neither resource reservation nor QoS is guaranteed. A number of fields are defined in the RTP header to specify what actions should be taken for a receiver when packets arrive. For example, the field *payload type* indicates the media type and the packet's compression way, such as audio or video

encoding. The fields *time stamp* and *sequence number* are used by the receiver to synchronize time-based media packets. RTP is accompanied with Real-time Control Protocol (RTCP), which periodically sends information about the quality of the packet transmission to all participants in the RTP session.

RTSP, initiated by Netscape, RealNetworks, and Columbia University, is an application-level protocol for controlling single or multiple real-time streams, such as video or audio delivered over the Internet (Muzio et al., 2002). Similar to HTTP, RTSP uses textual commands to control stream transmission. However, unlike HTTP, RTSP is a state-oriented protocol, by which an RTSP server and client need to maintain the state of the connection session labeled by a session identifier. It offers the functionality of VCR-like remote control so that the client may fast-forward, rewind, pause, or stop the media streams. Furthermore, in contrast to HTTP, both server and client can issue requests. For instance, the server may ask the client to connect to another media server for services. RTSP request/response messages in a session often are transmitted in a transport-level (either TCP or UDP) channel independent of the media channel. The underlying transport standards for an RTSP session to deliver media streams can be various, ranging from TCP, UDP, multicast UDP, or RTP.

SMIL-ENABLED MULTIMEDIA-STREAMING ASSETS

In SCORM, the mechanism of content packaging is designed for packing instruction contents and the navigation sequences. In each content package, the oranizations portion specifies the content structure and the behaviors of this content. The resources session describes the instruction materials needed in this content, the type and physical location of each physical file. To create an SMIL-enabled asset, one may define an unlaunchable asset, which refers to a physical filename ending with .smil (e.g., "intro.smil", as shown in Figure 2 and Table 1). In addition, a <metadata> element can be used to describe its attributes. The instruction unit specified in the package is composed of one SCO:R_S1, which, in turn, consists of five assets as R_L1, R_U1, R_U2, R_U3, and R_U4. In this case, the instruction is packed in the content package, named "MMC.pif", and the content of imsmanifest.xml in "MMC.pif" is shown in Table 1.

The intro.smil document defines the rendering sequence of text (para.txt), image (2122.gif), sound (s1.wav) objects and real-time media (in intro.smil) synchronously in different regions, respectively (see Table 2).

The physical resource for SCO R_S1 is the file SCO.html, which is shown in Table 3. It consists of two javascript files packed in the content package. This SCO communicates with LMS by invoking some functions like doLMSGetValue, and all functions called by SCO.html are defined in either APIWrapper.js or AUFunctions.js.

After importing the content, LMS will identify each element in the content package and place them in the appropriate positions in order to show them when learners launch this content through the runtime environment. Figure 3 illustrates how the content package is composed, the process of launching contents, and the communication with the LMS server and the RTP/RTSP server.

Each SCO must contain at least, one launchable asset (e.g., R_L1 in Table 1) that contains javascript files and is responsible for communicating with LMS. When the SCO communicates with LMS, the launchable asset is connected to APIAdapterApplet, which, in turn, communicates with LMS afterward. LMS invokes EJB objects, residing in the server, to access LMS DB with the data model that stores the information of SCOs tracked by different LMSs.

The SMIL document can be defined as an unlaunchable asset similar to the jpeg asset when a content package is compiled. Since an SMIL

Figure 2. Files tree structure packaged in "MMC.pif"

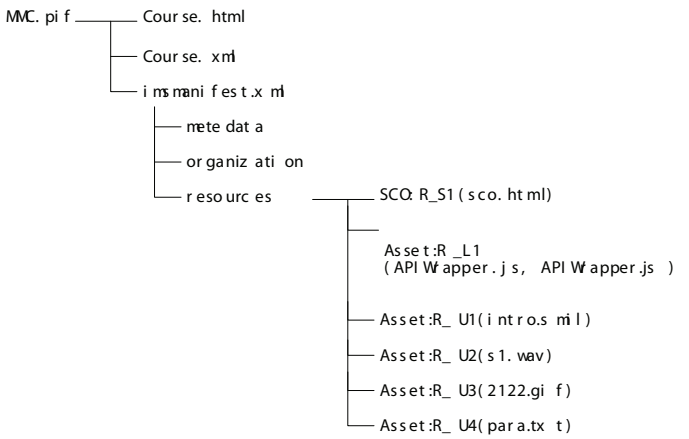


Table 1. Content package "MMC.pif" with a SMIL-enabled asset: imsmanifest.xml

```
.....

<organizations default="a0">
  <organization identifier="a0">
     <title>SMIL Navigation</title>
     <item identifier="S1" identifierref="R_S1" isvisible="true">
       <title>Introduce</title>
     </item>
     ....
  </organization>
</organizations>


<resources>
  <!--sco-->
  <resource identifier="R_S1" type="webcontent"
  adlcp:scormtype="sco" href="sml01/sco.html">
     <metadata>
       <schema>ADL SCORM</schema>
       <schemaversion>1.2 </schemaversion>
       <adlcp:location>sml01/sco.xml</adlcp:location>
```

document may be composed of several files, all files in the SMIL document must be packed as resources in the content package (see Table 1). Moreover, an SMIL document may contain real-time multimedia objects that refer to an RTP server. The RTP server sends a stream of bytes containing the media back to the client. The SMILAppletAdapter buffers a portion of

Table 1. Content package "MMC.pif" with a SMIL-enabled asset: imsmanifest.xml (cont.)

```
    </metadata>
    <file href="smil01/sco.html" />
    <dependency identifierref="R_L1" />
    <dependency identifierref="R_U1" />
    <dependency identifierref="R_U2" />
    <dependency identifierref="R_U3" />
    <dependency identifierref="R_U4" />
</resource>

 <!--launchable asset-->
<resource identifier="R_L1" adlcp:scormtype="asset"
 type="webcontent" xml:base="smil01\Scripts">
    <file href="AUFunctions.js" />
    <file href="APIWapper.js" />
</resource>

<!--unlanchable asset-->
<resource identifier="R_U1" type="webcontent"
 adlcp:scormtype="asset">
    <metadata>
      <schema>ADL SCORM</schema>
      <schemaversion>1.2 </schemaversion>
      <adlcp:location>smil01/resource/intro.xml</adlcp:location>
    </metadata>
    <file href="smil01/resource/intro.smil" />
</resource>
```

the media, which the client begins playing after a certain portion has been received. Additional media are continually buffered, providing users with an uninterrupted clip.

RENDERING OF SMIL-ENABLED MULTIMEDIA-STREAMING ASSETS

Since most Web browsers do not support the functionality of rendering SMIL documents, we develop a Java applet-based player by extending the SMIL player presented in Li and Chen (2001). The applet specification is embedded in the content, as shown in Table 4. When the instruction content is presented to a learner, the player automatically will be downloaded and installed onto the client's browser and the SMIL document rendered at the same time.

The SMIL player is built upon Java Media Framework (JMF) and Java-XML technologies to overcome the issues of interoperability and

Table 1. Content package "MMC.pif" with a SMIL-enabled asset: imsmanifest.xml (cont.)

```
<resource identifier="R_U2" type="webcontent"
 adlcp:scormtype="asset">
   <file href="smil01/resource/s1.wav" />
</resource>

<resource identifier="R_U3" type="webcontent"
 adlcp:scormtype="asset">
   <file href="smil01/resourc e/2122.gif" />
</resource>

<resource identifier="R_U4" type="webcontent"
 adlcp:scormtype="asset">
   <file href="smil01/resourc e/para.txt" />
</resource>

</resources>
.....
```

Table 2. The content of SMIL resource:intro.smil

```
<smil>
 <head>
   <meta name="noname" content="empty" skip-content="true" />
    <layout type="text/smil-basic-layout">
      <root-layout height="250" width="350" skip-content="true" />
      <region id="sound" height="198" width="21" left="17" top="8"
       z-index="1" fit="meet" skip-content="true" />
      <region id="video" height="197" width="298" left="36" top="8"
       z-index="2" fit="meet" skip-content="true" />
      <region id="sub" height="200" width="317" left="17" top="204"
       z-index="3" fit="meet" skip-content="true" />
      <region id="main" height="160" width="230" left="20" top="10"
       z-index="1" fit="fill"/>
   </layout>
 </head>
```

integration in distributed component computing, so that its constituents can be reused and sharable. The player contains three major components: XML Parser, Protocol Handler, and Render Engine. XML Parser, an event-oriented parser based on the Simple Api for XML (SAX) model, vali-

Table 2. The content of SMIL resource:intro.smil (cont.)

```
< body>
  < seq repeat="1" >
    <par endsync="last" repeat="1" >
     < audio id="audio-0" region="sound" src="s1.wav" dur="3s"
       repeat="1" fill="remove" />
     < img id="img-0" region="video" src="2122.gif" dur="10s"
       repeat="1" fill="remove" />
     < text id="text-0" region="sub" src="para.txt" dur="3s"
       repeat="1" fill="remove" begin="1s" />
     < rtp id="img-1" region="main" src="rtp://224.0.0.1:999/audio"
       dur="250s" fill="remove" />
    </par>
    <rtsp region="main" src="rtsp://163.18.16.220/ProgID=1036806870"
     dur="250s" fill="remove" />
  </seq>
</body>
</smil>
```

dates the specified SMIL document and retrieves the embedded elements and attributes. Protocol Handler invokes the procedures for handling specific protocols, such as HTTP/HTML, JPEG, WAV, and RTP media. Finally, Render Engine is responsible for the appropriate rendering of different media objects.

Since the SMIL-type resource is not supported by most LMSs, the approach of the applet-based player provides a lightweight, feasible solution for augmenting the functionality of LMSs. If the SMIL standard were eventually adopted by the SCORM society, the player could be embedded into APIAdapterApplet of LMS; this definitely will simplify the development task of SCORM contents.

SCORM-COMPLIANT LMS IN EJB

To tackle the interoperability issue of e-learning platforms, we developed the proposed SCORM-compliant SMIL-enabled multimedia streaming LMS on Java Enterprise Java Beans (EJB) environment (Jboss, 2003). The EJB environment, part of Java 2 Enterprise Edition (J2EE), is a technology for developing, assembling, deploying, and managing distributed component applications. It simplifies the handling of issues inherent in distributed applications by concealing such complexities as security, transaction handling, and database access, and enables component developers to focus on business logic. There are two types of enterprise Java beans: session and entity. Session beans are business objects designed to act as the interface between business logic and application logic, whereas entity beans are data objects. In the study, the session bean calls out to a single underlying entity bean to access the database model. The mechanisms of EJB, such as business objects, data objects, object pool, and transaction, are particularly suitable for developing the data model in SCORM RTE.

Table 3. Segments of sco.html

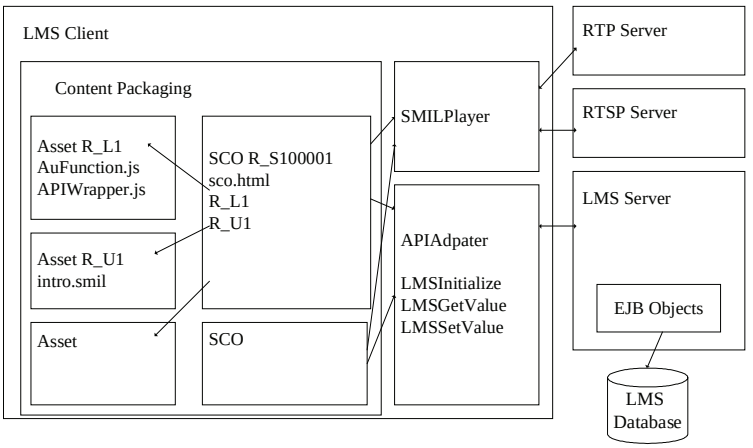
```
<HTML>
<HEAD>
  <SCRIPT LANGUAGE=JAVASCRIPT SRC="scripts/APIWapper.js"></SCRIPT>
  <SCRIPT LANGUAGE=JAVASCRIPT SRC="scripts/AUFunctions.js"></SCRIPT>
  <TITLE>SCORM introduce</TITLE>
</HEAD>

<BODY onLoad="loadPage()" onunload="return unloadPage()">
<SCRIPT language="javascript">
  var studentName="!";
  var lmsStudentName = doLMSGetValue( "cmi.core.student_name" );
  if(lmsStudentName!="")
     studentName=" " + lmsStudentName + "!";

  document.write( studentName + " enjoy SML content.<br>");
</SCRIPT>
<!--SML Player-->

</BODY>
</HTML>
```

Figure 3. The act of Content content Packaging packaging in LMS Clientclient



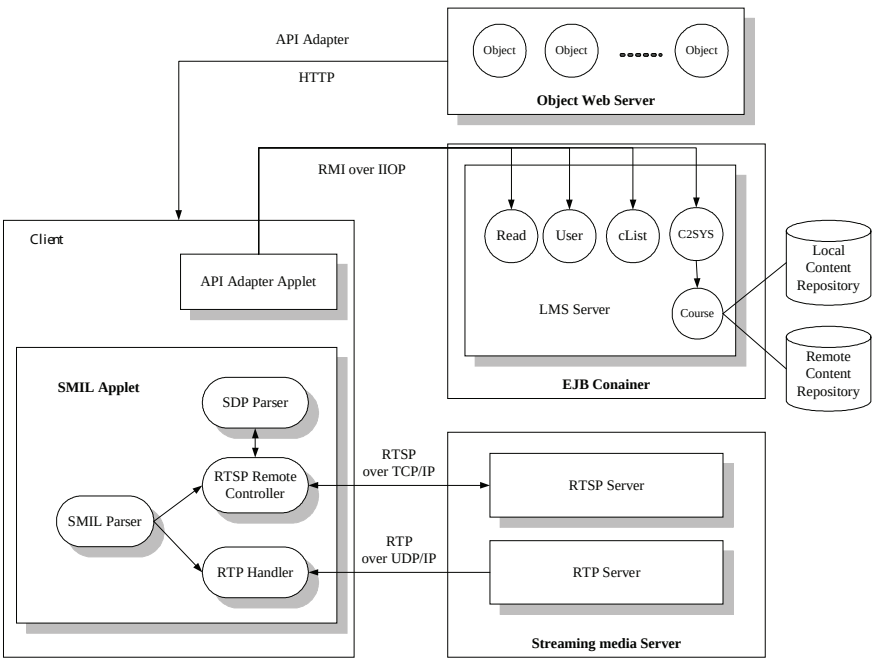
System Architecture

Figure 4 shows the system architecture of the proposed LMS developed in the EJB environment. First, the client has to download an API Adapter Applet from the_Object Web Server. An API Adapter Applet is an EJB client that is responsible for communicating with the LMS Server, so that a client can assess learning material in LMS. The LMS Server follows EJB's model, and there are five enterprise Java beans in LMS Server, which are discussed as follows.

Table 4. The specification of an applet-based SMIL player in the content:sco.html

```
<applet archive="sSmil.jar" code="SmilApplet.class"
 width=450 height=380>
   <PARAM NAME ="location" VALUE="./smil-intro/intro.smil">
   <PARAM NAME ="height"   VALUE="380" >
   <PARAM NAME ="width"    VALUE="450" >
</applet>
```

Figure 4. The architecture of the proposed LMS in EJB environment

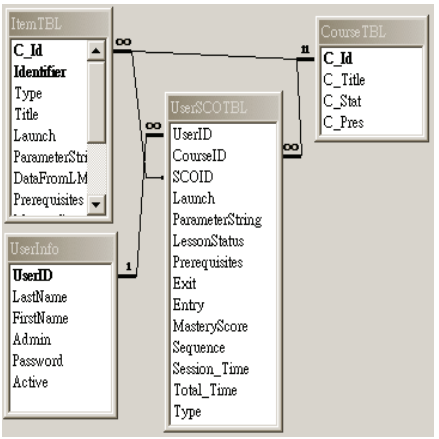


- **UserEJB:** UserEJB is an entity bean, that is in charge of confirming the legality of users and maintaining their information.
- **cListEJB:** cListEJB is an entity bean, that maintains the list of instructions available in LMS so that students can register as regular students or guests and choose any course they like to browse.
- **ReadEJB:** ReadEJB is an entity bean. The learning path of each student will be recorded in the LMS Server, if an instructor wants to keep track. When students are reading, taking tests, or setting preferences in an instruction, the ReadEJB bean will track and record such behavior in the data model defined in SCORM's RTE in the background. This bean plays the most important role in the proposed LMS.
- **C2SYSEJB:** C2SYSEJB is a session bean that allows teachers or system administrators to import courses or update course contents. In detail, this EJB object is responsible for unzipping the content package (a PIF file) of instructions, analyzing the MANIFEST, and allocating all physical files in appropriate locations in a content repository. Moreover, this session bean is the interface between EJB client and CourseEJB.

- **CourseEJB:** CourseEJB is an entity bean that endorses the C2SYSEJB bean to access the database of content repository. In implementation, this entity bean is wrapped by CourseEJB in a session bean, so one can access all business functionality without the overhead of polling the entity bean.

In order to support manipulating multimedia streams, the SMIL Parser parses SMIL-enabled assets and extracts embedded multimedia objects.

Figure 5. The entity-relationship of the LMS database



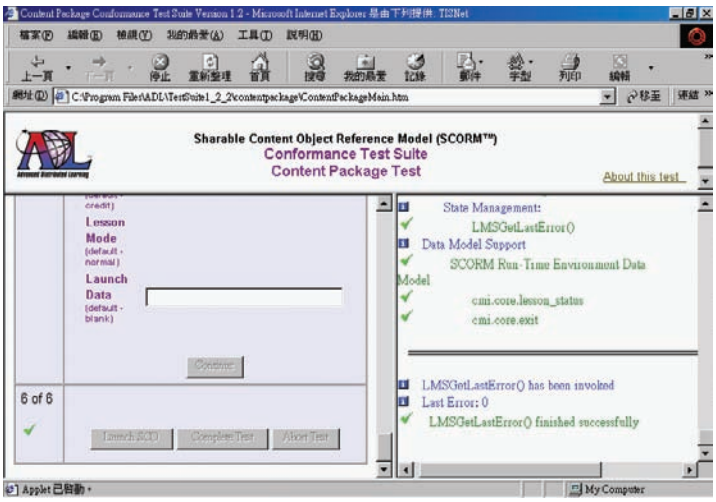
If the extracted object is an RTP object, the RTP Handler will be invoked to receive RTP media streams, synchronize audio and video stream, and render them in the client screen. On the other hand, when the client obtains an RTSP object, an RTSP Remote Controller also is called. Initially, the RTSP Server returns a Session Description Protocol (SDP) message, which describes the information of the RTP media to be transmitted. After parsing the SDP data by SDP Parser, RTP Handler receives and consumes the RTP streams. During playback, the RTSP Remote Controller handles the RTSP protocol between client and the RTSP Server, so that the client may issue play, pause, fast forward, rewind, or stop commands to control the streams.

Schema of Database

The database in the LMS Server is the place to store all detailed data, such as the individual information, specifications of instructions, learning paths, and so forth. The main schema of the database is depicted in Figure 5.

There are four major tables defined in the database. First, table UserInfo collects all the individual information. Specifications of each instruction are maintained in CourseTBL as an

Figure 6. Test SMIL-enabled content with Test Suite



unit. ItemTBL stores the detailed information on every learning resource, such as asset, SCO, and aggregation. Finally, UserSCOTBL stores the students' learning behaviors, such as student name, score.max, core.exit, and so forth. All fields are designed adhering to the data model in SCORM RTE. From the schema, one notes that fields and relations are designed according to attributes of objects and roles of objects in LMS, respectively.

Figure 7. The import the content package containing SMIL-enabled assets

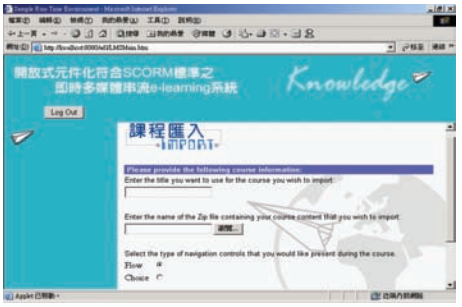
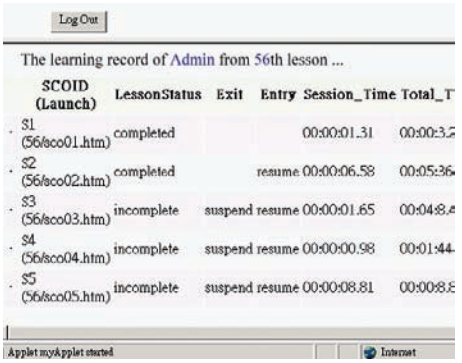


Figure 8. The presentation of the RTP in SMIL-enabled SCORM content on learner's browser

Figure 9. The presentation of the RTSP in SMIL-enabled SCORM content on learner's browser

Figure 10. The learning path tracked



EXPERIMENT AND DEMONSTRATIONS

The proposed SMIL-enabled multimedia streaming e-learning system has been experimented on JBoss-3.0.3_Tomcat-4.1.12 EJB server (Jboss, 2003). To demonstrate the effectiveness of the system, the SMIL-enabled SCORM content package (shown in Table 1) is imported to the proposed LMS, but only after the conformity with the SCORM standard of SMIL-enabled content is validated. SCORM Version 1.2 Conformance Test Suite Version 1.2.2 provides a simple tool for self-testing. The test suite provides conformance test from four options:

- LMS Run-Time Environment Conformance Test
- SCO Run-Time Environment Conformance Test
- Meta-Data Conformance Test
- Content Package Conformance Test

Figure 6 shows the steps of the successful tests of the proposed SMIL-enabled instructions

against the conformances of SCO Run-Time Environment, Meta-Data, and Content Package.

Subsequently, the content can be imported to the LMS server, as depicted in Figure 7. We choose the type of navigation control as *choice*, and this finishes the registration procedure. Figure 8 is the snapshot of browsing SMIL-enabled RTP content, and Figure 9 illustrates an RTSP multimedia object instead. The tiny difference between the two snapshots is that there is a VCR-like control bar below the movie in Figure 9.

From the snapshots demonstrated, SMIL instructions are rendered on the learner's browser appropriately, whenever the instructions contain raw media, such as sounds, graphs, texts, and so forth, or the read-time types RTP/RTSP. One notes that the tree structure of the course content is listed clearly in the left-down frame as well, and learning can proceed on the learner's demand. Figure 10 illustrates the learning path of some students recorded by ReadEJB. With this functionality, the teacher is granted the ability to monitor or query the learning path and the students' progress in the course, and thus, appropriate measures can be taken to improve the instruction quality.

CONCLUSION AND FUTURE WORK

This paper proposes an SMIL-enabled asset model that allows W3C's SMIL documents to be embedded in the instructions. The asset model is enhanced further by incorporating multimedia streaming objects. Realizing and rendering such assets on learners' Web browsers, a Java applet-based SMIL player and LMS are needed. To meet the reusability of software development, we integrate the components of the player developed in our previous work, SAX for parsing SMIL documents and JMF for playing the RTP/RTSP media. To handle the interoperability issue of LMS, we developed the proposed LMS in the overwhelming Java EJB component-based computing environment. The initial experimental results show an encouraging achievement. With SMIL ability and streaming multimedia objects of assets, instruction contents could be more vivid and interactive. The proposed asset model is adherent to the SCORM standard; thus, it can be sharable, reusable, accessible, and durable. As far as the references surveyed, our work presented is the very pioneering study in the literature toward integrating SCORM, SMIL, and multimedia streaming in e-learning. Based on the lessons learned in this study, the next research direction is to enhance the proposed LMS by incorporating directory services, such as Lightweight Directory Access Protocol (LDP), so that remote SMIL-enabled assets can be interoperable and sharable. Another objective could be aimed toward developing an efficient and effective way of maintaining these assets using ontology engineering for supporting resource and knowledge sharing.

ACKNOWLEDGMENT

This work was supported in part by the National Science Council, Taiwan under NSC91-2520-S-327-001 and NSC92-2524-S-006-004.

REFERENCES

AICC. (2003). Aviation Industry CBT Committee. Online *http://www.aicc. org/*

Anido, L., Nistal, M. L., et al. (2001a). A distributed object computing approach to e-learning. *Proceedings of the Third International Symposium on Distributed Objects and Applications*, (pp. 260-269).

Anido, L., Nistal, M. L., et al. (2001b). A standards-driven open architecture for learning systems. *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, (pp. 3-4).

Anido, L., Nistal, M. L., et al. (2001c). CORBA-based runtime environments for standardized distributed learning architectures. *Proceedings of the Informing Science Conference*, June 19-22, Krakow, Poland.

ARIADNE. (2004). Alliance for Remote Instructional and Authoring and Distribution Networks for Europe. Online *http://www.ariadne-eu.org*

IMS. (2003). IMS content packaging specification version 1.1.2. Online *http://www.imsglobal.org/*

Ismail, J. (2002). The design of an e-learning system beyond the hype. *The Internet and Higher Education, 4*, 329-336.

Jboss. (2003). Jboss™ The Porfessional Professional Open Source Company. Online *http://www.jboss.org/index. html*

Li, S.-T. (2002). A Web-aware interoperable data mining system. *Expert Systems with Applications, 22*, 135-146.

Li, S. -T., & Chen, H.-C. (2001). An architecture-neutral approach to Web-based synchronized multimedia presentations with RTP video streams. *Proceedings of The Seventh International Conference on Distributed Multimedia Systems (DMS2001),* Tamkang University, Taipei, Taiwan, September, 26-28.

Li, S.-T. (2002). A Web-aware interoperable data mining system. *Expert Systems with Applications, 22*, 135-146.

LMML. (2004). Learning material markup language. Online *http://www.lmml.de/*

LTSC. (2004). IEEE Learning Technology Standards Committee. Online *http://ltsc.ieee.org/wg12/*

Muzio, J.A, Heins, T., & Mundell, R. (2002). Experiences with reusable e-learning objects from theory to practice. *The Internet and Higher Education, 5,* 21-34.

SCORM. (2003). Sharable content object reference model version 1.2. Online *http://www.adlnet.org/*

SMIL. (2004). Synchronized multimedia. Online *http://www.w3.org/AudioVideo*

ULF. (2002). ULF (Universal Learning Format). Online *http://www.saba.com*

This work was previously published in International Journal of Distance Education Technologies, edited by S.-K. Chang and T.K. Shih, pp. 48-64, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.7
Feature-Based Multimedia Semantics:
Representation for Instructional Multimedia Design

Michael May
LearningLab DTU, Technical University of Denmark, Denmark

ABSTRACT

The aim of this chapter is to sketch a semantic taxonomy of representational forms (or "sign types" in the terminology of semiotics) relevant for the compositional analysis, design, and indexing of multimodal multimedia, specifically, in the domain of instructional design of learning objects, and to contribute to current attempts to bridge the "semantic gap" between the technical and spatio-temporal description of media objects and the contextual perception and interpretation by users of their content.

INTRODUCTION

In the present chapter, it is suggested that descriptions of *graphical and multimedia content*, as it is known from multimedia databases, multimedia programming, and graphic design, should be extended to include *taxonomic information about the representational forms* (or "sign types" in the terminology of semiotics) combined in multimedia presentations.

From the point of view of design, there is a fundamental *rhetorical* problem in deciding which *combination of media types and representational forms* ("sign types", "modalities of expression") would be adequate for a particular multimedia presentation within a particular context-of-use. This rhetorical problem is necessarily present in all forms of information, interface, and interaction design, although it is not the whole story of designing usable and adequate presentations, since there are many criteria of usability as well as adequacy of presentations (i.e., presentations that are relevant, intelligible, efficient, consistent, aesthetic, etc.) and many levels of design that have to be considered in the design or automated gen-

eration of presentations (functional specification, spatial layout, temporal organisation, forms of interaction, forms of cognitive support for users, contextual relations to activities and situations of use, etc.). In this chapter, we will only address the rhetorical problem of "media and modality" and specifically the problem of taxonomy, that is, the need for some kind of classification of media types and representational forms to support a conceptual understanding of the design space of available types and its operational specification in particular design contexts (where the example given here is multimedia content specification for instructional design).

TAXONOMIES OF MEDIA AND REPRESENTATIONAL FORMS

Taxonomies in Graphic Design

One might think that taxonomic issues in graphic design have been settled for a long time, since the representation and layout of simple graphical objects like two-dimensional documents with "text" and "illustrations" appear simple compared to the full complexity of computer-based animation, video, gesture, and haptics. This appearance of simplicity is an illusion, however, and the problems of representation and layout of multimodal documents have not yet been fully solved (Bateman, Delin, & Henschel, 2002; Bateman, Kamps, Kleinz, & Reichenberger, 2001). It is important, from the point of view presented in the present chapter, to note that *multimedia* communication as well as *multimodality* (of representations) does not originate with modern computer technology or electronic audio-visual technologies like television. We know multimodal representations from the printed graphical media in the form of combinations of language, images, and diagrams, and before the invention of printing, from combinations of writing and drawing within the graphical media. The fundamental origin of multimedia communication and multimodality, however, is coextensive with the origin of language, since the embodied language of human speech combining multiple media (speech, gestures) and multiple forms of representation (natural language discourse, intonation patterns, schematic "conversational gestures", facial expression, etc.) is an advanced form of multimodal communication (Quek, McNeil, Bryll, Duncan, Ma, Kirbas, McCullough, & Ansari, 2002) that in some ways can play the role of a natural ideal for designed technically-mediated communication.

The taxonomic problem of identifying and combining different types of media and representational forms (or "modalities") have been addressed in design theories based on the printed graphical media, but it was intensified with the advent of computer-based multimedia systems. It will be useful to start the present analysis of problems of taxonomy with a few early attempts to classify graphical objects.

With a few exceptions, the focus on taxonomic issues within *graphical design* has been on representational forms derived from examples of "business graphics", and it has been restricted to static graphical media objects. In the *scientific visualisation* community, on the other hand, there has also been some interest in taxonomic issues, but with a focus on images, maps, and graphs used for data visualisation within simulations and other forms of dynamic graphics. Early work in scientific visualisation was driven by presentations of examples of new applications and graphic algorithms, whereas the conceptual foundation of the domain was neglected. Recent work in the domain of *information visualisation*, common to business graphics and scientific visualisation, has attempted to present the area as based on knowledge of perception (Ware, 2000). This is an important step forward, but what is still missing is an understanding of the syntactic and semantic features of graphical expressions, and how these features are expressed in other media (acoustic, haptic, gestic).

With the increasing importance of the Internet and Web-based presentations, a new quest for taxonomies has arisen in studies of automated presentation agents and multimodal document design. It appears that, even from the point of view of document design restricted to static two-dimensional documents including text, pictures, and diagrams, there is still no consensus on a systematic description of multimodal documents (Bateman, Delin, & Henschel, 2002).

A taxonomy of graphics in design was proposed by Michael Twyman (1979). "Graphical languages" is here analysed along two dimensions: *configuration* and *mode of symbolisation*. The focus of Twyman was on static graphics and the role of eye movements in reading graphics. Configuration thus refers to different forms of linear, branched, matrix-like, or non-linear reading of graphical forms, whereas the "mode of symbolisation" refers to the verbal/numerical, verbal and pictorial, pure pictorial, or schematic organisation of the content. Ordinary text is usually configured typographically to be read in a linear fashion, whereas diagrams with branching nodes have a branching linear configuration, and pictures are examined in a non-linear fashion. Although we are presented with a classification of the "modes of symbolisation", there is no analysis of the semantic content as such. The classification is inconsistent in considering combined forms such as "verbal and pictorial" to be a class next to simple forms such as "pictorial" from which it must be derived. Twyman refers to the distinction between symbolic and iconic in semiotics, but does not find it relevant for the classification, and he does not utilise his own conception of "graphical languages". Even though the concept of a language implies syntactic and semantic structures, he only considers "pictorial" and "verbal" forms as aggregated and unanalysed wholes.

Another example of classification problems with "graphics" is seen in the work of Lohse, Walker, Biolsi, and Rueter (1991) and Lohse, Biolsi, Walker, and Rueter (1994). The methodology is here quite different; through a cluster analysis of individual subjective classifications of a series of examples of graphics, they arrive at the following distinctions: graphs and tables, network charts, diagrams, maps, and icons. The authors try to interpret these classes along two dimensions, but they are given no real analysis. The five classes seem to range from symbolic (graphs and tables) to more "analogue" forms (maps and diagrams) in one dimension, and from explicit and "expanded" forms of expression (network charts) to more implicit and "condensed" forms of expression (icons) in another.

The work of Lohse, Walker, et al. (1991) and Lohse, Biolsi, et al. (1994) should be understood as referring to "folk taxonomies" derived from clustering of perceived differences among representations. This kind of taxonomy could be important for studies of individual styles of reasoning and self-generated external representations (Cox, 1999), but less important for systematic models of design choices in multimedia design. "Perceived differences" will reflect peoples' attitudes towards representations rather than their intrinsic characteristics, and folk taxonomies are not consistent classifications: Sometimes people will group representations together because they have the same use context, sometimes because they are expressed in the same media, and sometimes because they "look alike" (but disregarding their semantic and syntactic properties).

Web-based techniques for presentation based on HTML and XML have enforced a clearer separation of document content from its layout as well as from different types of meta-data associated with documents, and this is reflected in recent theories of the structure and layout of multimodal documents. According to Bateman, Delin, and Henschel (2002) and Bateman, Kamps, Kleinz, and Reichenberger (2001), it will be necessary to distinguish five types of document structure: the *content structure* of the information communicated in a document, the *rhetorical structure* of the relations between content elements, the

layout structure of the presented document, the *navigation structure* of the supported ways in which the document can be accessed, and the *linguistic structure* of the language used to present the document. The question of media types is *implicit* in this typology since only graphical media documents are considered in these typographic theories, whereas the question of representational forms (or "sign types") is hidden in what is called "the linguistic structure". For our purpose, the linguistic structure should be generalised to the *semiotic structure* of multimodal multimedia documents or "objects".

Media Types

In early approaches to multimedia the concept of "multimodality" was constructed from an engineering point of view as referring to combinations of technologies. According to Gourdol, Nigay, Salber, and Coutaz (1992) and Nigay and Coutaz (1993), for example, a modality is associated with a physical hardware device. A multimodal system is accordingly defined as a particular configuration of hardware, that is, as a system that supports many "modalities" for input and output. The taxonomy that can be constructed from this point of view will be restricted to dimensions of the technical control of hardware systems such as the support given by these modalities with respect to temporal execution (sequential or parallel) and the combined or independent nature of commands with regard to the modalities. This taxonomy of "systems" does not consider the representational issues involved in multimedia, but from the perspective of users of applied multimedia systems, the important design decisions cannot be based on a characterisation of the physical devices and media, since it is precisely the representational and cognitive properties of the application that will determine what kind of support users can expect in their situated activity involving the application. An example is instructional design of multimedia applications for higher learning, where Recker, Ram, Shikano, and Stasko (1995) have stated the importance of what they call the "cognitive media types": Design of hypermedia and multimedia for learning should not be based directly on properties of the physical media, but on properties of the information presentation that are "cognitively relevant" to the learning objectives.

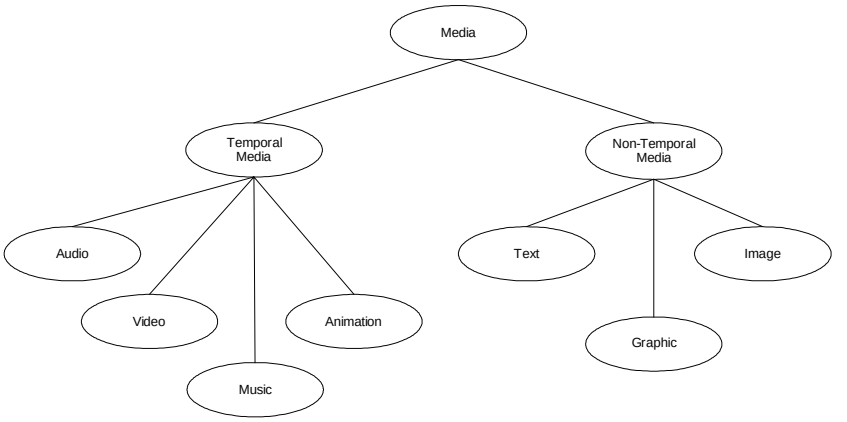
An important step forward towards a classification of media types was taken with the object-oriented approach to multimedia programming, but even with the focus on a conceptual foundation for multimedia (Gibbs & Tsichritzis, 1995), the software abstractions constructed to represent media types were mainly a reflection of the data types and digital file formats available at the time. The concept of media type introduced the idea of a template for distinguishing the representational from the operational aspects. With a template like

```
Media type <name>
    Representation
            <aspects of representation>
    Operations
            <categories of operation>
```

Gibbs and Tsichritzis (1995) identified a set of important media types as well as their associated forms of representation and supported operations. The actual classification of media types represented by media classes was, however, both incomplete and inconsistent.

The suggested media types are incomplete from a systematic point of view, where we would want to include haptic and gestic media types. Another problem is the inherent inconsistencies of the classification scheme. The distinction between temporal and non-temporal media appears clear cut, but on a closer analysis all so-called non-temporal media have temporal aspects. Text can indeed be statically displayed, but it could also be presented repetitively (blinking), sequentially (one word or noun phrase at a time displayed in a

Figure 1. Media classes as suggested by Gibbs and Tsichritzis (1995); the diagram has been redrawn to make the assumption of the non-temporal media classes explicit in the diagram



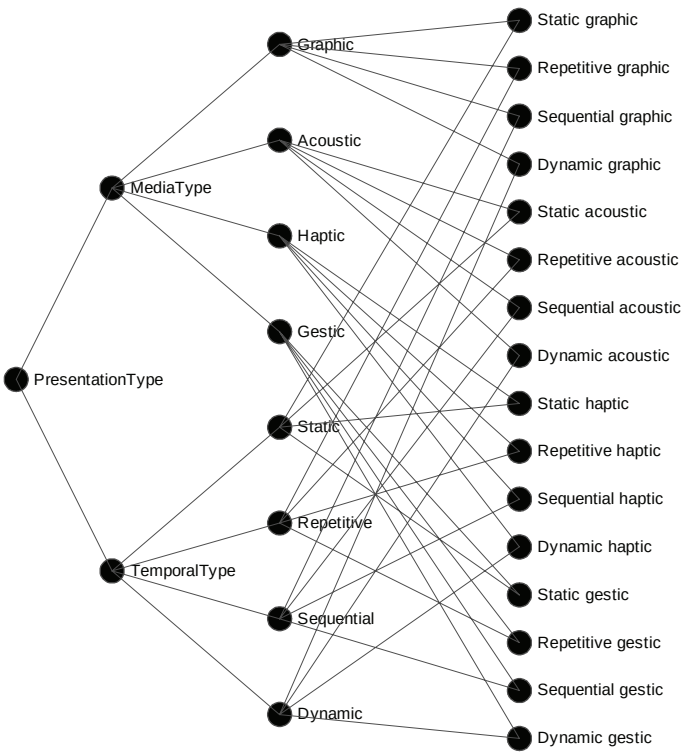
linear fashion as in the "ticker tape display"), or dynamically as in auto-scrolling text on a Web page. The same argument can be made for images and object-oriented graphics, that is, they are not inherently static, but can be presented in different temporal forms. The media type called "animation" now appears to be a derivation of repetitive, sequential, or dynamic versions of other media types. It could be an improvement to differentiate the temporal properties of presentations from the specification of their media type (just as the interactivity of presentations is seen as a separate issue), and a classification scheme like the one constructed in Figure 2 could be defended.

From the classification presented in Figure 2, a series of questions will have to be answered, though. How can, for instance, the acoustic media be ascribed a static form? Although the acoustic media is inherently temporal at a physical level, it does make sense from a perceptual point of view to define a static acoustic media type with reference to sound (in the form of noise or un-modulated tones) that is constant within the duration of the presentation. After considering limit cases like static sound, the same temporal distinctions can be applied to all media resulting in a classification scheme (Figure 2) based on the systematic formation of all logical combinations. The temporal distinctions are static, repetitive, sequential, and dynamic (Bernsen, 1993, 1994).

With regard to the difference between "animation" and "video" suggested by the traditional classification scheme (Figure 1), this distinction breaks down when we consider modern techniques of animated computer graphics and virtual reality in the 3D computer games or in mixed-media film, where animated computer-generated characters and human actors can appear within the same media object. Although animated graphics and video have different technical foundations and origins, there is no longer any fundamental difference between them as systems of communication. The referential difference between "virtual" graphics and "real" photographic presentations have important cognitive and epistemic ramifications, but it does not determine the media type (i.e., graphic/photographic and animation/video are not distinctions in media type).

Finally "animation" is a mixed concept for another important reason: It includes the traditional graphical forms of communication utilised in animated drawings (sequential graphics) as well

Figure 2. A systematic classification of presentation types as derived from media types and four temporal types, as they can be determined beofre the further articulation of representational forms (and before considering spatial and temporal layout issues)



as the embodied movement-based communication realised in facial, hand, and full-body gestures. As channels of communication, we need to distinguish the *gestic media* ("gestures") from the graphical media, although presentations in both channels are visually perceived.

There is a more serious problem with the temporal specification of media types, though: Although temporal specifications in some cases seem to be fundamental to the internal properties essential to media types (as in the case of static graphics versus animation and video), this does not resolve the issue of how this internal temporality is articulated within the temporal organisation of the presentation as a whole. In accordance with the different levels of structure introduced for multimodal documents, we need to separate different issues of temporality in multimedia objects.

Temporal aspects of multimedia objects can appear as important for the type specification of the chosen presentation media for a particular representational content, for instance, whether a graph or a diagram within an instructional document is animated or not. With present-day techniques for multimedia multimodal presenta-

tions such as Java applets for interactive Internet documents or dynamically-constructed graphs within a Maple worksheet for doing mathematical calculations, the difference between static, sequentially-animated, or fully-dynamic graphs or diagrams turns out to be of minor importance at the level of media types. Static, animated, or dynamic versions of the same object will often be available as part of the temporal layout of the instructional document or otherwise put directly under the interactive control of the user. Consequently, distinctions like static, repetitive, sequential, and dynamic should be located at the level of temporal layout and interactivity and not at the level of specification of media types (as an internal property of the type).

Gibbs and Tsichritzis (1995) are aware of another classification problem that arise with the combination of "audio" and "video": We would usually expect "video" to have some sound overlay, but since both audio and video can exist in a "pure" form, we should classify them according to these simple non-combined types and consider audio-visual media types as combinations of simple types (like video, audio, speech, and music).

A more difficult problem of classification, however, is associated with the inconsistent analysis of representational forms supported by different media. Speech is not considered as a type except as collapsed into the audio type, partly because it, in contrast to music represented in the midi encoding format, did not have a separate encoding at the time (cf. the recent VoiceXML format), and partly because there is no thorough analysis of the relation between media types and the representational forms they support.

In early approaches to multimedia content, many obvious relations that would be relevant for content management and for flexible display of multimedia content were almost "hidden" by the classification scheme. A good example is natural language, where the close relation between *text* and *speech* is lost if graphical and acoustic language presentations are only considered as presentations in two different media, and not also as representations within a single *representational form* (natural language). Similarly, the classification scheme hides the fact, well known to any graphical designer, that text is in fact a kind of graphical presentation that inherits many of the operations supported by other forms of object-oriented graphics: operations on shape, size, colour, shading, texture, and orientation (cf. the design space of fonts). These inconsistencies can be resolved, if we abstract the concept of media from specific technologies, programming languages, and their data types, in favour of a perception-based approach to different media, where media types are derived from human sensory abilities in combination with elaborated channels of communication. From this point of view, we only have four different physical presentation media and channels of communication subsumed under computerized control today: the graphic media that derives from visual perception, the acoustic media that derives from auditory perception, the haptic media that derives from tactile and kinaesthetic perception, and the gestic (or "gestural") media that derives from visual perception. The gestic and the graphic should be considered as separate channels of communication because the gestic media is based on the temporal dimension of movement in itself, rather than on its potential for producing graphic traces of movement (as in drawing on some interface).

Sign Types: Core Semantics and Emergent Semantics

From the point of view presented here, many problems of classification arise from the confusion of the physical media of presentation and communication with the representational forms ("sign types"). Natural language is a representational form that can present linguistic content across different media in the form of graphical language (text), in the form of acoustic language (speech), in the form of haptic language (Braille-

embossed text), or in the form of encoded gestures (sign languages). A useful starting point for understanding representational forms is, in fact, to consider them as representational "modalities", not to be confused with sensory modalities, that are invariant across different media (Stenning, Inder, & Neilson, 1995). In their analysis of media and representational forms (which they call "modalities"), Stenning, Inder, and Neilson focused on the invariant differences: The difference between diagrams and language, for instance, is invariant across the tactile and the graphical media. Where the media type is derived from how an object of presentation is communicated and perceived, the representational form (sign type) is derived from how it is interpreted. (the intended interpretation within a context). Braille text as well as graphical text is interpreted as language, whereas graphic and tactile diagrams are interpreted as diagrams. A semiotic and cognitive theory of information presentation is needed to explain this semantic invariance of the intended interpretation, by which representational forms are constituted. It can be claimed, however, that language is a unique representational system in providing this invariance across physical forms, that is, across the phonological and gestic system of expression and across different graphical writing systems, due to the differential nature of linguistic entities. Everything in language is differences, as it was proclaimed by de Saussure.

According to Stenning and Inder (1995), graphical representation systems are different because they rely on a less abstract mapping of relations between the represented domain and relations in the representation (within a presentation media). This should explain why there are graphic and tactile diagrams but apparently no acoustic equivalent, because the transient nature of acoustic media object does not allow diagrammatic relations to be presented and inspected. It is possible, however, to modify the idea of invariant properties as constitutive of the representational forms. According to the modified view, we will stipulate that only a reduced set of core properties are invariant across media and this core is what constitutes the representational form. On top of these core properties, each media affords and provides cognitive support for an extended set of properties and operations that are meaningful within the representational form, although they are not supported in all media.

From this point of view we could in fact have acoustic diagrams, but only in the very limited sense made available for the core properties of diagrams that are supported across all media. Acoustic diagrams would, therefore, be highly schematic with few details, but they could be constructed from external representations corresponding to the embodied image schemata described in cognitive semantics (Johnson, 1987; Lakoff, 1987). Acoustic diagrams would be based on image schematic asymmetries like left/right, up/down, centre/periphery, front/back with simple objects, relations, and events plotted against this background. An example could be the representation in sound of the relative position and movement of enemy aircrafts represented in the cockpit of a fighter aircraft through directional sound: The image schematic background for the acoustic diagram would in many cases be implicit (cf. the discussion of implicitness and expressiveness in Mackinlay & Genesereth, 1985), that is, it would not be explicitly represented but provided through the interpretation by the pilot according to the phenomena of distributed cognition (Zhang, 1996). Another example could be the use of directional 3D sound to indicate location of escape routes on passenger ships, where the traditional static graphic emergency signs could be misleading (in cases of fire blocking some escape routes) and difficult to see (due to smoke) (May, 2004).

It follows from this conceptualisation of media types and representational forms, that a distinction has to be made between the core semantics of the representational forms (sign types) and the extended semantic properties of the well-known prototypes associated with each representational

form. The acoustic maps mentioned above are good examples. They are indeed maps, a diagrammatic representational form, in the restricted sense of the core semantics of the unimodal map, according to which a map is interpreted as a diagram supporting the representation of spatial localization of objects and/or events relative to a reference object. Prototypical maps as we know them from everyday life or professional work (city maps, road maps, sea charts, etc.) are graphical objects with a higher complexity resulting from (a) emergent properties of the expression of maps within the graphical media (such as support for graphically-based inferences about metrical properties like size and distance) and resulting from (b) properties inherited from other representational forms combined with the map object. Most maps are in fact complex multimodal objects because they combine useful features of many unimodal representational forms in order to achieve the usefulness of the prototypical map object. A unimodal city map without overlay of symbols in the form of street names, for instance, would not be very useful. In the case of city maps the combination of types can be understood as a spatially coordinated layering of different forms: a network chart (metro lines) on top of the basic map object (the street map), and symbols (street and metro names) as yet another layer. This conceptual layering of different representational forms is utilised constructively in geographical information systems (GIS), where each layer can be addressed and modified separately.

Towards a Semantics of Multimedia: The Semiotic Turn

A semiotic turn in the analysis of multimedia has been initiated by the Amsterdam multimedia group at CWI. Nack and Hardman (2002) refer to the work of Chatman on film theory (Chatman, 1978) in order to set up the fundamental conceptual aspects of multimedia: the form and substance of multimedia content and the form and substance of multimedia expression. It should be noted that these distinctions of form/substance and content/expression does not originate with film theory, but with the constitution of structural linguistics according to Ferdinand de Saussure (1993) and Louis Hjelmslev (1961). Saussure and Hjelmslev both stipulated a place for a future science of semiotics (or "semiology") derived from the linguistic study of language, but with a focus on the syntactic and semantic properties of other "sign systems". Building on Saussure's distinction between the plane of expression (the "signifiers") and the plane of content (the "signified") in language and on his foundation of linguistics on the differential forms in language as a system, as opposed to its physical "substance" (acoustic, graphic) and its psychological "substance" (ideas, concepts, etc.), Hjelmslev specified these four planes in language (Hjelmslev, 1961).

The four planes can be used to specify different levels of description for multimedia units as shown. A similar description is given by Nack and Hardman (2002), although with a different interpretation of the levels. A more complex semiotic framework for the analysis of multimedia based on structural linguistics in its further development by A. J. Greimas was given by Stockinger (1993).

Note that the standard developed for multimedia content description interfaces for audiovisual media, better known as the ISO MPEG-7 standard, specifies multimedia units at the level of their substance of expression only, and the MPEG standard thus leaves the problem of mapping high-level semantics of media objects unsolved (Nack & Hardman, 2002). Even on its own level as a standard for the substance of expression, there has been some criticism of MPEG-7 for not being extensible and modular enough in its description definition language: It is criticised for its closed set of descriptors, for its non-modular language, and for not being sufficiently object-oriented in its data model (Troncy & Carrive, 2004).

Figure 3. The four plans in language as a system accoring to Saussure/Hjelmslev tradition

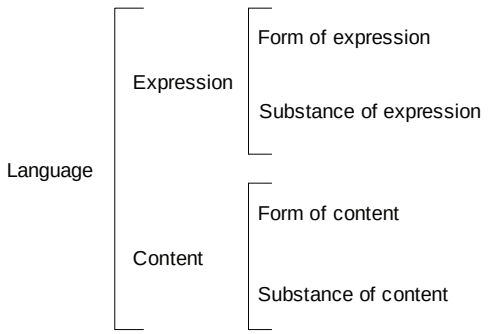
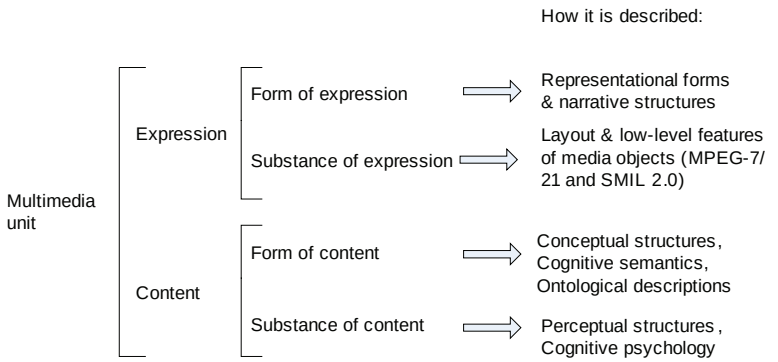


Figure 4. Four fundamental levels of description of multimedia units (not considering the pragmatic and social aspects of the use of multimedia in a specific context-of-use)



On the level of the form of expression, our focus here is on the basic representational forms rather than on higher-level semantic structures of narration and interaction, and it is comparable to the three-dimensional TOMUS model presented in Purchase and Naumann (2001). The TOMUS model is based on a differentiation between sign types, syntactic structures, and sensory modalities, the latter corresponding to *media* in the present study.

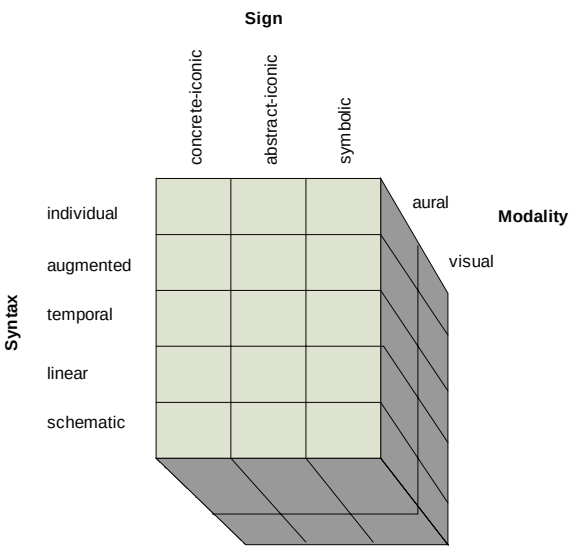
Some taxonomies of multimedia reproduce the difficulties encountered by taxonomies of graphics. An example is the ambiguous classification of media types into text, sound, graphics, motion, and multimedia as suggested by Heller & Martin (1995) and Heller, Martin, Haneef, and Gievska-Krliu (2001). "Motion" is here used to refer to video and animation, which seems like a confusion of form and substance of dynamic graphics. The haptic and gestic media are not considered. The relation between text and graphics is left unanalysed, and the inconsistency in not considering text as a form of graphics is not discovered. The model, however, introduces a second dimension of expression of media types covering three nominal types going from "elaboration" over "representation" to "abstraction". In the case of graphics, this expression dimension is exempli-

fied by the difference between photographs and images at the "elaboration" end, blueprints and schematics as the intermediate form of "representation", and icons at the "abstraction" end. The expression dimension of Heller and Martin (1995) is close to the dimension of the sign in the TOMUS model of Purchase and Naumann (2001), with its subdivision of sign types into concrete-iconic, abstract-iconic, and symbolic.

The iconic-symbolic dimension of sign types have been derived from the semiotic analysis of signs in the work of C. S. Peirce. Where Saussure and later Hjelmslev conceived "semiology" as an extension of linguistics that would study the quasi-linguistic properties of other "sign systems", Peirce developed his more general conception of a "semiotic" from his analysis of logic. For Peirce the sign is not a dyadic relation between expression and content, but a triadic relation between a physical representation (in his terminology, the "representamen"), and an interpretation (the "interpretant") of this representation as referring to an object in some respect. Within this triadic relation, it is the representation-object aspect which is categorised as being iconic, indexical, or symbolic by Peirce. The causal relation implied by the indexical category can be considered as a separate issue from the dimension of iconic – symbolic. A subdivision of the iconic – symbolic dimension was also used as a foundation of the sign typology presented in (May, 1993, 2001; May & Andersen, 2001), with the main categories being image, map, graph, conceptual diagram, language, and symbol. The image and the map category here corresponds to the concrete-iconic sign type in the taxonomy of Purchase and Naumann (2001), the diagrammatic forms of representation – graph and conceptual diagram – correspond to the abstract-iconic sign type, and language and symbol to the symbolic type. The conception of the "diagrammatic" in C. S. Peirce is somewhat broader (Figure 6), since it would include maps in the concrete-iconic end and natural as well as mathematical languages in the symbolic end.

With regard to media, the TOMUS model only distinguishes the aural and the visual senses

Figure 5. The TOMUS model of sign types (i.e., representational forms), syntax, and (sensory) modalities (i.e., media) according to Purchase and Naumann (2001).



supporting the acoustic and the graphic channel of communication. It is, however, an important feature of the model that all combinations of sensory modalities (or rather media) and sign types (representational forms) can be considered in a systematic way. We will not go into the discussion of the syntactic dimension of the model here.

The taxonomy suggested in May (2001) and May and Andersen (2001) relies on a set of fundamental principles, which makes it well suited to support a more formal approach. An initial version of this taxonomy was developed in May (1993) within the GRACE Esprit Basic Research project in collaboration with N. O. Bernsen, who developed his own version (Bernsen, 1993, 1994). It is suggested that:

- There is a limited number of media types derived from sensory modalities and forms of communication relevant for computer-based interaction;
- There is a limited number of possible uncombined representational forms ("unimodal sign types");
- It is possible to give a feature-based description of their semantic properties;
- Some of these properties are invariant across media types whereas other properties and operational possibilities are emergent with the expression of signs within a media;
- Invariant properties (the "core" semantics) are inherited to syntactic combinations of representational forms whereas their emergent properties are constrained by the specific combinations of media types and representational forms.

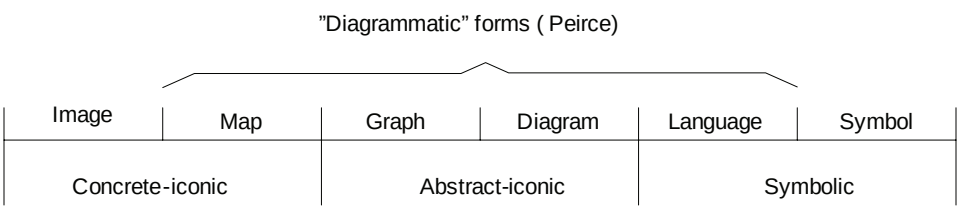
It follows that we cannot expect to find any kind of consistent hierarchical classification of prototypical multimedia objects or even graphical objects, because they will have to include complex objects with overlapping properties. The consistent alternative is to develop a feature-based classification, where we track the inheritance and combination of significant properties of sign types and media types. This will, of course, require that we can define the core semantics of the suggested representational forms and describe them in a way that will be relevant to the actual interpretation and use of complex media objects, where different media and representational forms are combined.

FEATURE-BASED MULTIMEDIA SEMANTICS

Feature-Structures and Multimedia Content

An early suggestion of a feature-based semantics of graphics was given by Alan Manning in a paper on "technical graphics" (Manning, 1989). Manning stipulated four types of technical graphics

Figure 6. Conceptual diagram to indicate the relation between the sign types used in the TOMUS model compared with the taxonomy of representational forms



abstracting from their terminological variants and their apparent similarities and differences: chart, diagram, graph, and table. According to his analysis, these types can be distinguished by logical combination of two features: display of one unit (written as –u) or several units (+u) and the representation of one property (-p) or several properties (+p). Each graphical type in this simple semantic system can thus be described by a small set of feature structures:

chart: [-p, -u] graph: [-p, +u]
diagram: [+p, -u] table: [+p, +u]

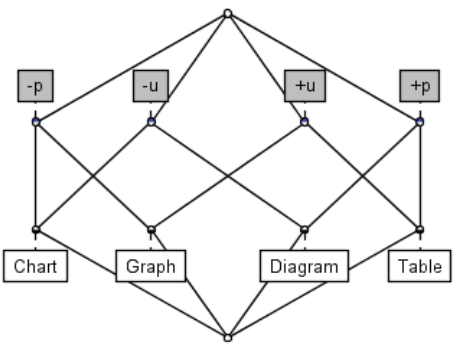
A simple pie chart only displays one unit (some totality) and it only represents one property (fractions of the total), whereas a graph like a bar graph has several units (represented by individual bars) and again only one property (the amount represented by the height of each bar). A diagram, however, only represents one unit (some object), but displays several properties of the represented object, whereas a table can represent many units and display several properties of each represented object. Our interest here is in the formal approach using feature structures rather than the actual classification.

From a formal point of view, the feature structure approach corresponds to setting up a lattice of the logical combinations of features. This approach to taxonomy of conceptual structures has been formalised in Formal Concept Analysis (Ganter & Wille, 1999), where features ("attributes") and the concepts they specify (the "objects") are related in a matrix called a formal context. A formal context C:= (G, M, I) is defined as two sets G (from German "Gegenstände", Objects) and M (from German "Merkmale", attributes) with a relation I between G and M. The elements of G are the objects and the elements of M are the attributes or features of the context. From the formal context, all possible combinations of formal concepts can be generated. A formal concept is a pair (a, b) where a belongs to the set of objects G and b belongs to the set of attributes M. In the small example derived from Manning (1989), an example of a formal concept would be the specification of charts as ({Chart}, {-p,-u}). The full list of formal concepts for the context is shown in Figure 7 below together with the Hasse diagram of the corresponding lattice. To construct a lattice, we have to add two points: the "top" element corresponding to the empty set of objects and the union of all attributes and the "bottom" element corresponding to the full list of objects and the empty set of attributes.

Let us return for a moment to the analysis of the core semantics of a representational form

Figure 7. Formal concepts and lattice generated for the graphical types (objects) and the features (attributes) given by the example discussed earlier; the lattice drawing has been made using the Java-based Concept Explorer program "ConExp 1.2".

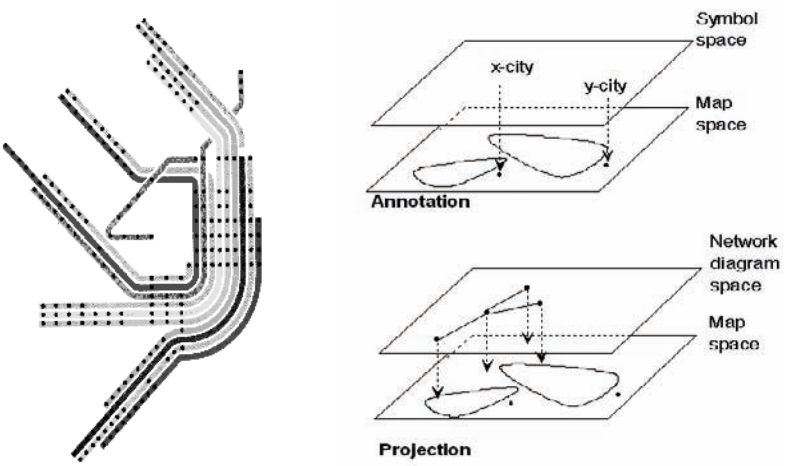
({}, {-p,-u,+p,+u}) = Top
({Chart}, {-p,-u})
({Graph}, {-p,+u})
({Diagram}, {-u,+p})
({Chart,Diagram}, {-u})
({Table}, {+p,+u})
({Chart,Graph}, {-p})
({Diagram,Table}, {+p})
({Graph,Table}, {+u})
({Chart,Diagram,Graph,Table}, {}) = Bottom



(across all media) as opposed to its emergent semantics articulated within the graphical media. The diagrams used to locate city subway stations are sometimes referred to as "maps", although they are in fact network charts with features quite distinct from the features of maps. Subway charts will often be juxtaposed with timetables, which belong to yet another type of representation, because the combination of tables with information about the time schedule of trains and the network diagram (representing the connectivity and sequence of train stations) gives effective support for navigational decisions. Both representations are, however, virtually useless without the addition of a third representational form: symbolic representations for the denotation of different train stations. The representational form of symbols is also needed to establish a coherent reference between the annotation to the network chart and the content of the table structure. The annotated network charts can furthermore be projected on an underlying map object giving rise to the familiar but complex prototype known as a "city map". City maps are complex multimodal representations that are not fully coherent in their inherent semantics, because the supported operations and the semantic properties of network charts do not fully agree with those of maps. Since network charts focus on the representation of topological properties like the connectivity of nodes and connecting lines in the network and not on its metric properties, the connecting lines can be deformed and distances between nodes can be changed: As long as the connectivity of the network is not changed, it will have the same interpretation (i.e., it will represent the same network). These operations are, however, not supported in a map and there is, therefore, a potential conflict of interpretation in maps with network overlay. In city maps the nodes representing subways stations have to be fixed to their correct location on the map, whereas the shape of connecting railway lines can be idealised and thus not true to the underlying map. This is acceptable because the geometric properties of railway lines are unimportant to passengers. On the other hand, city maps could support false inferences about these properties as seen on the map.

The core semantics of a representational form will necessarily be unimodal even though most concrete practical examples of external

Figure 8. Artificially reconstructed "unimodal" network chart (derived from a subway map) (to the left) and the (re-) construction by annotation and projection of a multimodal map object by spatially coordinated layering of different representational spaces (to the right)



representation will be multimodal. We can artificially construct a unimodal map by "reverse engineering" of a city subway diagram like the one shown below to the left (an early version of the Copenhagen subway system), but it is virtually useless because the labelling of stations have been removed.

We can utilise the Formal Concept Analysis (FCA) for graphical types and generalise it to cover the full range of multimodal multimedia objects. FCA have already been extended to cover constructive aspects of diagram design (Kamps, 1999) and multimodal document layout (Bateman, Kamps, Kleinz, & Reichenberger, 2001), but what has not yet been fully recognised is the potential benefits of extending the description of objects with semiotic metadata about the representational forms articulated and combined within different media. This type of information would not only be relevant for the constructive generation of multimodal multimedia documents, but also for the design of flexible distributed documents, for instance, in supporting changes in representational form in moving Web-based documents to PDA interfaces. The issue of flexibility here is not just a question of how to deal with layout constraints imposed by different devices, but also a question of designing support for adaptation or tailoring through transformations of representational scales (Johansen & May, in press), media types as well as representational forms (sign types).

Consider again maps and network charts as two different representational forms. We can express what this difference means, in terms of the semantic features and supported operations associated with each form within different media, by setting up the formal context for these simple "unimodal" objects. The following example is just given as a sketch of what such an analysis could look like, and this simplified formal context only includes five objects and four attributes (semantic features and supported operations):

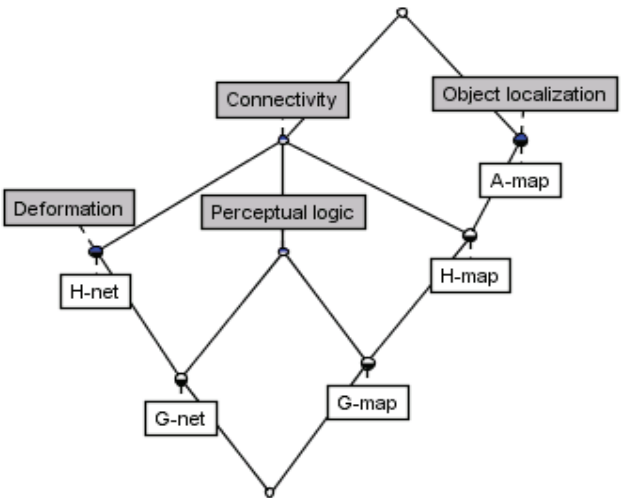
- **Connectivity:** Does the type require connectivity to be preserved?
- **Perceptual logic:** Does the type support "direct perception" of logical relations in a spatial structure (cf. Lakoff, 1990; May, 1999)?
- **Deformation:** Does the type support and allow deformation?
- **Object localization:** Does the type support localzation of objects relative to a background object?

It is implied by this analysis that the well-known graphical maps, although they are prototypical maps, do not exemplify the *core semantics* of maps. It is in fact the apparently peripheral case of the acoustic map (A-map in Figure 10) that exemplifies the *core* semantics of maps, because it is the attributes of acoustic maps (in the example here reduced to the feature "Object localization") that is shared by all maps. Graphical maps, of course, also have this feature, but in addition they have other emergent features

Figure 9. Table indicating a part of the formal context specifying the map and the network chart types articulated in different media

C	onnectivity	Perceptual logic	Deformation	Object localization
Graphical map (G-map)	+	+		+
Haptic map (H-map)	+			+
Acoustic map (A-map)				+
Graphical net (G-net)	+	+	+	
Haptic net (H-net)	+		+	

Figure 10. A Hasse-diagram of the lattice corresponding to the formal concepts specifying maps and network charts according to the simplified formal context sketched in Figure 9



derived from the graphical media (Connectivity and Perceptual logic).

We can visualise this through the inheritance hierarchy embedded within the lattice as shown in Figure 11. The core feature of maps (object localization) as exhibited by acoustic maps is inherited to haptic maps and graphical maps.

Lattice structures are well-suited to express feature structures, because they can express inheritance relations as well as systematic combinations of types. We can even use them to generate possible combinations of representational forms in cases where we might not know a prototypical example in advance, that is, we can use lattices to explore the design space of all possible type combinations and their expression in different media —given a full set of features necessary to distinguish the formal concepts involved.

The concept of a lattice is based on the concept of an ordering of types through the binary relation ≤. Defining this relation on a set, we obtain a partially ordered set (sometimes just called a partial order), if for all elements of the set, the following is true:

$x \leq x$ (Reflexivity)
$x \leq y \ \& \ y \leq x$ implies $x = y$ (Anti-symmetry)
$x \leq y \ \& \ y \leq z$ implies $x \leq z$ (Transitivity)

If every pair in the set has both a "supremum" (a least upper bound) and an "infimum" (a greatest lower bound) within the set, we have a lattice ordering <L, ≤>. From any lattice ordering <L, ≤> we can construct a lattice algebra <L,, > with two binary operations called join () and meet () by defining x y = sup{x,y} and x y = inf {x,y}. This equivalence of the order theoretic and the algebraic definition of a lattice is important, because it means that we can construct operations on a lattice given the ordering of its elements (Davey & Priestley, 1990). This is what is utilised in formal concept analysis (Ganter & Wille, 1999).

Figure 11. The lattice of Figure 10 used to show the inheritance of the "core feature" of maps (object localisation) from the acoustic map type; the object localisation feature is shared by all maps.

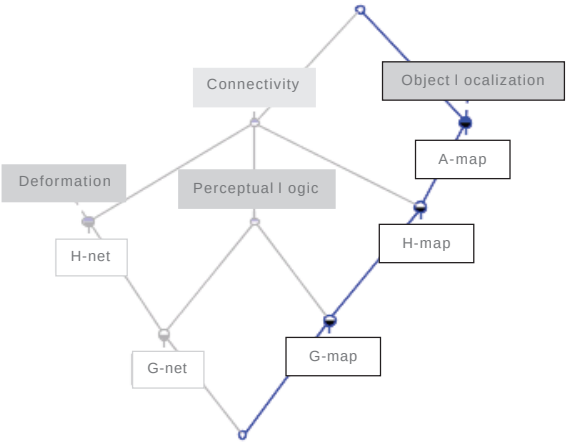
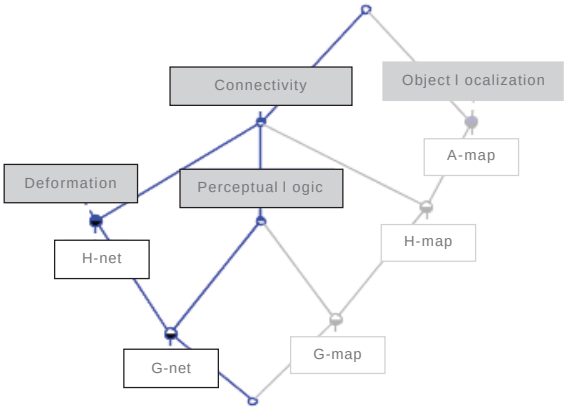


Figure 12. The lattice of Figure 10 used to show the specification of graphical nets (G-net) as the lattice "meet" of Deformation (a core semantic feature of network charts) and Perceptual logic (an emergent feature of the graphical media); by inheritance graphical nets are also characterized by Connectivity (a shared feature of the graphical and haptic media).



FEATURE-BASED TAXONOMY FOR INSTRUCTIONAL DESIGN

Related Approaches to Multimedia Semantics

The extension of feature-based approaches to multimedia semantics has followed different tracks, since the first attempts to specify features of graphical objects. One track has followed the work on automated presentation of information (Mackinlay, 1986) and another track have analysed perceptual and representational scales for relational information displays (Petersen & May, in press; Zhang, 1996). A third track can be identified around the problem of indexing im-

ages as well as indexing and sequencing video for databases (Dorai & Venkatesh, 2001, 2003), but this work has mainly dealt with low-level features, that is, features of the graphical substance of expression. Even when explicitly addressing the "semantic gap" between the physical substance of recorded film and the semantic issues of content management systems, the kind of semantics that is proposed is based on feature extraction from low-level data, although guided by film grammar. The focus on feature extraction is grounded in the practical interest in automatic segmentation and indexing of video. A fourth track is linked to the attempt to define semantic content for multimedia and hypermedia on "the semantic Web", but the standards developed for Web-based multimedia and hypermedia also have its main focus on the substance of expression and to a lesser extend on the form of expression. A good example is the advanced XML-based spatial and temporal layout facilities for media objects realised in the Synchronized Multimedia Integration Language (SMIL 2.0) (Bulterman, 2001).

From SMIL Encoding of Multimedia to Learning Object Metadata

The SMIL specification language for multimedia on the Web and for mobile devices can in many ways be seen as the realisation of the potential inherent in object-oriented multimedia with regard to the substance of expression. With SMIL 2.0, authors can specify the temporal behavior of individual elements of a presentation, their spatial layout, the hypermedia structure of the distributed objects included in the presentation, and its overall compositional structure.

The basic syntax of SMIL presentations is very similar to HTML syntax and is easy to read and edit. For a simple example, look at the following presentations that could be part of a multimedia instruction in how to play chess. Here a chess game has been visualised through a video recording of a particular game. The SMIL object is played back through the RealPlayer, but could also be embedded on a Web page. The first piece of code simply presents the recording of the board as captured in the file "board.rm".

A more advanced example is shown below, where the space within the player window is subdivided into four regions, one of which is used to display a textual comment synchronised with the moves of the particular chess game. The other three regions are used to display different visual perspectives on the game: the "black" player, the board, and the "white" player. In parallel with the synchronised graphical channels, an audio channel is added to present additional speech comments on the game.

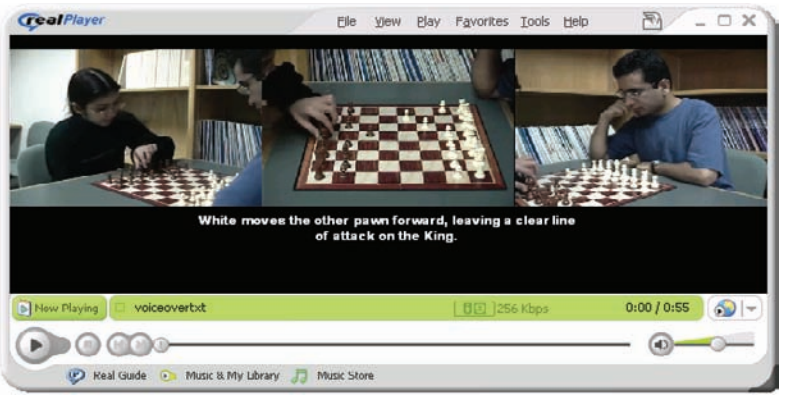
The chess example above has been produced in a project on digital video in educational research and it is accordingly not meant as an instructional object for teaching or learning how to play chess. We can, however, use it here as a point of departure for discussing the relation between the encoding of multimedia documents and the metadata required for their use in instructional design of learning objects.

A learning object is a modularized self-contained unit of instructional material that has been given a metadata description of its own content and use. Ideally these instructional materials are available in digital electronic form through a learning object repository (shared over the Internet, for example). As described by David Wiley, "learning objects are elements of a new type of computer-based instruction grounded in the object-oriented paradigm of computer science. Object-orientation highly values the creation of components (called "objects") that can be reused … in multiple contexts. This is the fundamental idea behind learning objects: instructional designers can build small (relative to the size of an entire course) instructional components that can be reused a number of times in different learning contexts." (Wiley, 2000, chap. 3, p. 1). There are however, as Wiley also points out, important

Figure 13. SMIL code use to present a recorded movie sequence

```
<smil>
  <body>
    <video src="board.rm" clip-begin="16s" clip-end="27s"/>
  </body>
</smil>
```

Figure 14. The RealPlayer used to play back a SMIL presentation of a chess game using four graphical regions as well as an audio channel for speech comments; the corresponding SMIL code is shown in Figure 15; this Public Domain example is provided by the Cognitive Development Lab of the University of Illinois at Urbana-Champaign



theoretical problems to be addressed with regard to the granularity of these objects (their minimal "size" within the instructional material of a whole course) and with regard to semantic and cognitive constraints (what makes sense, what gives cognitive support) as well as pragmatic constrains on compositionality (what has practical value within a context-of-use). We need to go beyond a naïve "Lego"-conception of learning objects.

The chess video example could be considered as a learning object provided that we associate it with relevant metadata describing its intended use within an instructional setting. It is in fact already a learning object within its original context, where it is used as an illustration of how to use SMIL for editing and presenting digital video. The very re-use of the chess example in the present context illustrates both the potential of learning objects and an inherent weakness: By completely changing the context and reusing the object with a purpose, for which it was not designed (chess instruction), it will be difficult to determine in advance, if the object is really useful and adequate in the new context. In any case, we will need detailed metadata describing the learning object and its intended use, but we cannot expect instructional designers to anticipate all possible context of reuse. If learning objects are to be retrieved and reused, they will therefore have to be designed with some level of flexibility and a combination of detailed information about the intended context-of-use (to support identifica-

Figure 15. SMIL code to present the RealPlayer presentation through an external URL (http://www.psych.uiuc.edu/~kmiller/smil/examples/voiceover.smil)

```
<smil>
<head>
  <layout>
   <root-layout width="1080" height="350"/>
   <region id="video_left" width="360" height="240" left="0" top="0"/>
   <region id="video_center" width="360" height="240" left="360" top="0"/>
   <region id="video_right" width="360" height="240" left="720" top="0"/>
   <region id="text_subtitle" width="560" height="100" left="260" top="250"/>
  </layout>
 </head>
 <body>
   <par dur="55s">
    <video src="black.rm" begin="5s" clip-begin="1.09s" region="video_left"/>
    <video src="board.rm" begin="5s" clip-begin="0s" region="video_center"/>
    <video src="white.rm" begin="5s" clip-begin="1.10s" region="video_right"/>
    <textstream src="text.rt" region="text_subtitle"/>
    <audio src="voicetrack.rm"/>
   </par>
</body>
</smil>
```

tion of similar learning situations) and abstract compositional descriptions of their form and content independent of any specific use (to support transformations of the object and its transfer to other learning situations). Before discussing these demands on metadata, let us stay with the fictive example of the chess learning object.

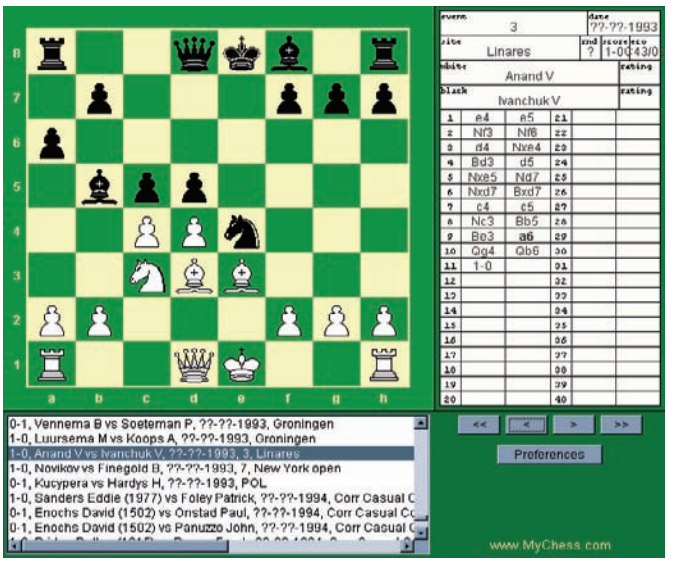
As a first observation, it should be noted that video playback only provides a limited level of interactivity. A learner would very soon in the learning process need to be actively engaged in playing chess (i.e., using a chess simulator or playing chess with a human opponent), rather than watching a game passively through recorded sequences of game play. On the other hand, even this simple example illustrates the learning potential of using multiple representations within multimedia learning objects. The chess example uses the representational form of language expressed as graphical text (for comments in the subtitle region) and as acoustic speech (for synchronised "voice over") in combination with dynamic graphics (the video). Combining a graphical display of the game with voice could be an effective support for basic learning according to the cognitive theory of multimedia learning proposed by Mayer (2001). According to Mayer's so-called "modality principle", students will learn better when the language part of a multimedia document is presented in the acoustic media (i.e., as voice) rather than in the graphical media (i.e., as text), because this use of voice for comments within the multimedia presentation will be less demanding on cognitive processing (since the visual-graphical channel is already used for static graphics, animation, or video). This principle has also been referred to as "dual coding theory", that is, referring to the utilisation of the capacity of visual as well as auditory working memory.

There is another important issue, however, that is not addressed adequately by the cognitive theory of multimedia learning. The "dual coding"

theory stays within a media-specific dichotomy between the visual and the auditory channel of communication, and representational forms are more or less reduced to "pictures" and "words". Although Mayer (2001) uses different diagrams in his empirical investigations, they are treated as "pictures" at the theoretical level. This makes it difficult to address didactic aspects of learning with different forms of representation. In our small example with the hypothetical learning object for multimedia support of chess learning (Figure 14), it is obvious that chess learning beyond the initial stages of getting acquainted with the context of chess playing would benefit from relying on "abstract-iconic" representations of the relational structure of the game and its strategic situations rather than on "concrete iconic" representations like images (animated or not). In other words, we need the familiar chess diagrams to support game analysis, and this requires that we abstract from the irrelevant features of the concrete situation of game play (unless our focus is the psychology of game players rather than the game itself).

In learning chess, we would in fact abstract diagrams representing the strategic situations from observing a video, but the mental work required to do so would be less demanding if these abstractions were supported by a sequence of external diagrams. This is an important educational implication of the variable cognitive support provided by different external representations according to the theory of distribution between external and internal (cognitive) representations (Petersen & May, in press; Zhang, 1996)). Image type representations will be less effective and less adequate as cognitive support for chess learning and understanding compared to diagrams (including their algebraic notation) for two different reasons: Images will necessarily qua images represent features that are irrelevant to the relational structure of the game (demanding us to abstract from them) and the symbolic structure of chess will

Figure 16. The MyChess Java applet illustrates a more powerful form of playback, where chess diagrams are used in combination with their algebraic notation to support analysis of the game (here Anand - Ivanchuk in Linares, 1993). The applet also presents a database from which the current game is selected (source: http://www.mychess.com)



require us to see in the image "something more" than is directly visible, that is, the diagrammatic form (May, 1999). This is why the external diagram and its algebraic equivalent are more effective, that is, less cognitively demanding.

Professional chess players will require more interactive features and more functionality to support further learning than can be realised with playback of recorded games. With the full simulation provided by a game engine, the expert will not only use a program like Fritz to play against a virtual opponent, but also to perform dynamic thought experiments on the external diagrammatic representations in order to explore the game space and evaluate different strategies. This exploration corresponds to what the founding father of logical semiotics, C. S. Peirce, called "diagrammatic reasoning" (Peirce, 1906).

Extending Learning Object Metadata

As shown in the previous example, it is generally not enough to know which media are utilised by some learning objects in order to evaluate their usefulness and their cognitive support for specific activities and contexts-of-use. We also need to know details about the representational forms used by learning objects. This type of information has not yet been included in the design of metadata for learning object repositories. If we again depart from a SMIL example of a learning object, but this time one that has been give a metadata description, we can take the SMIL presentation of a lecture in "Membranes and Cell Surfaces" from a BioScience course given at University of Queensland.

The metadata given is primarily formal (author, version, department, institution, etc.), deictic (date, place, presenter) and superficial content descriptions (subject, description). An important information about the level of learning that the learning object is assumed to address is given by the statement of the audience being "First Year Biology Students" at a university, but it is not stated what the specific learning objectives for the lesson is (what are students supposed to be able to do after using the object or after taking the course?), or indeed what form of learning and instruction is supported by the lesson? On inspecting the online lesson, it turns out to be a recorded audio and video sequence of a traditional lecture, with added notes and diagrams in a separate region. This type of presentation can be presented with a SMIL 2.0 player or embedded within a website. Alternatively a similar presentation can be made with Microsoft Power Point Producer (available for Office 2003). In any case there is no metadata standard for describing the representational forms used by learning objects and what kind of cognitive support this gives for the intended activity within a context-of-use.

Learning objects have been given several competing XML-based standards or specifications in the form of the IEEE Learning Object Metadata (LOM), the Instructional Management System (IMS) Metadata, and the Sharable Content Object Reference Model (SCORM), to mention just three different models. Although these standards and specifications are supposed to be neutral to different contexts-of-use, there have been many objections towards their implementation in higher education. The SCORM standard, for example, promoted by the Advanced Distributed Learning initiative (ADL), sponsored by the U.S. Department of Defense, has a bias toward distributed learning situations with single learners engaged in self-paced and self-directed learning (Friesen, 2004). This is well suited for instructional design of individual Web-based training, but less so for the pedagogical use of learning objects within group-based and class-based teaching and learning at the university level.

The metadata specified by the standards focus on formal document information, technical requirements, superficial content descriptions (like keywords), and general educational information (like interactivity level, resource type, duration of use), but generally neglect more specific infor-

Figure 17. Example of metadata given for a SMIL learning object presenting a Bioscience lecture; the information shown here is from the Australian DSTC repository (Source: http://maenad.dstc.edu.au/demos/splash/search.html#browse)

Subject:	Structure and properties of biological membranes
Description:	This is a lecture about the properties of membranes and cell surfaces
Presenter:	Dr Susan Hamilton (University of Queensland)
Date:	08-22 on the 22-08-2000
Place:	University of Queensland Saint Lucia
Audience:	First Year Biology Students at UQ
Duration:	0:18:52
Subject:	BL114 (Semester One): Introduction to Biological Science
Course:	B.Sci: Bachelor of Science
Department:	Biological Sciences
Institution:	University of Queensland
Editor:	Suzanne Little
Date.Modified:	15-11-2001
Version:	1.0
Source Presentations:	biosciences1

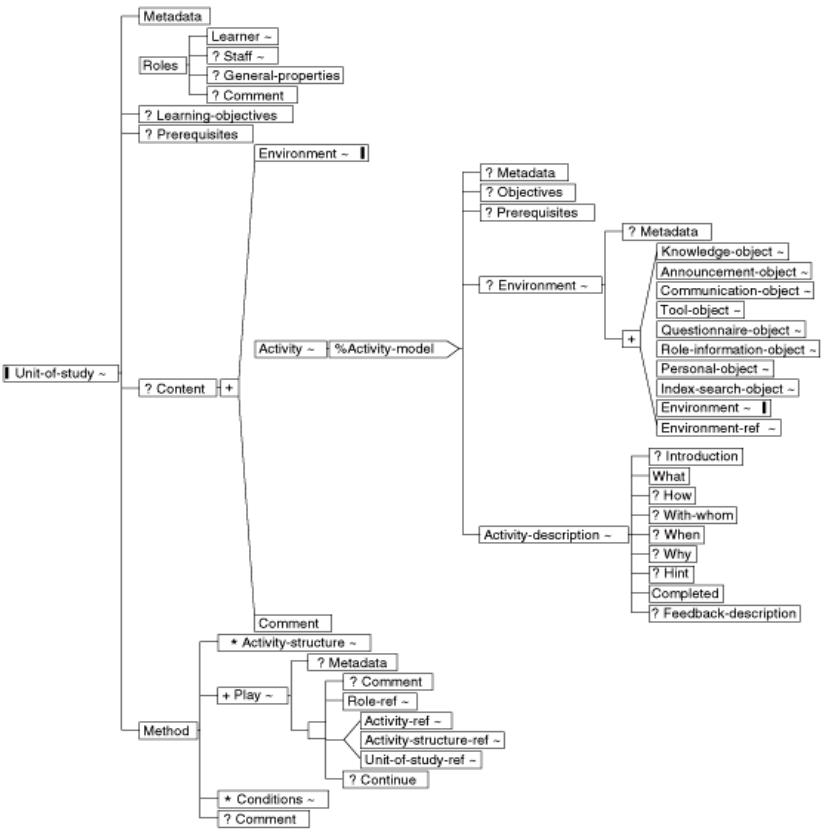
mation about the content. Content specifications should include more detailed information about the scientific domain to which the object refers, the prior knowledge required by the learner to benefit from its use, the learning objectives for the use of the object, and its intended scenarios of use. This type of information is included within the Dutch proposal for a more elaborated Educational Modelling Language (EML), which goes beyond learning objects in order to model the whole activity of teaching and learning, the methods used, and the different roles of the social agents involved (Koper, 2001; Koper & van Es, 2004). EML is proposed as "a major building block of the educational Semantic Web" (Anderson & Petrinjak, 2004, p. 289).

This extension of the modelling perspective from the objects of instructional design, that is, teaching materials, to include the whole activity of teaching and learning, might however be counter-productive and impractical. The perspective of learning objects should not be to formalise the whole activity of teaching and learning, but to package electronically-available learning objects with metadata relevant for teaching and learning, and to embed them in object repositories for future reuse. They should, however, be tagged with some of the information provided with EML, but only as tools for reuse, not as a part of a codification of the whole activity.

There are also alternative conceptions of learning objects like the Reusable Learning Assets (RLA) proposed within the Open Course initiative (http://opencourse.org), which are defined as open source, non-proprietary learning objects.

What is lacking in all projects, however, even in the detailed modelling of "units of study" in EML, is the relevant metadata information about the media types and representational forms utilised by the learning object. This information is relevant because there is evidence that variance and integration of representational forms are important for student learning, even if there has been some debate over the specific "modality effects" of media on learning. Early experimental work on the "effects of media" on learning did

Figure 18. EML specification of a unit of study



not produce consistent findings (Hede, 2002), but conflicting conceptions of learning have also confused the issue. In the "effects of media" debate there was originally a conception of learning as transmission of information, and teaching was accordingly seen as a simple matter of communication, whereas learning in modern conceptions involves constructive conceptual activity by students. Even in modern studies it is, however, not always clear whether the "modality effects" claimed for multimedia learning is due to the duality of audio and visual coding, that is, a consequence of the use of acoustic and graphic media types, or due to the multiplicity of different representational forms such as language, diagrams, graphs, and images.

THE EXAMPLE OF GRAPH COMPREHENSION

The theoretical critique of Mayer's cognitive theory of multimedia learning (CTML) has been backed up by empirical studies. Schnotz and Bannert (2003) found that adding graphical representations to text presentations did not always result in improved understanding. In some cases the "pictures" could actually have negative effects by interfering with the construction of mental models. They conclude that the "dual coding theory" of multimedia learning does not take into account the effects that different forms of visualisation will have on the construction of mental models from the presentation in the process of comprehension. In another study it was found

that the "affordances" of graphical representations for improving conceptual understanding should be seen in the context of specific scientific domains and also as dependant on familiarity with the specific "iconic sign system" from which the representations are derived (De Westelinck, Valcke, De Craene, & Kirschner, 2005). It seems, for instance, that students in the social sciences are less familiar with graphs and conceptual diagrams than students of natural sciences, and they will therefore not have the same benefits from these abstract-iconic forms. Learners can only benefit from representations if they understand the semantics of the representational forms to which they belong (Cox, 1999).

A striking example is the problems of graph comprehension reported in many studies of science learning in high school as well as in higher learning. An early example is the studies of students' understanding of kinematics and problems of associating graphs with their physical interpretation (Beichner, 1994; McDermott, Rosenquist, & van Zee, 1987). A recent review of the literature on graph comprehension (Shah & Hoeffner, 2002) supports these early findings and lists four implications for teaching graphical literacy: (1) graphical literacy skills should be taught explicitly (and in the context of different scientific domains), (2) activities where students have to translate between different types of representations will improve their understanding, (3) students need to pay attention to and be trained in the relation between graphs and the models they express, and (4) students should be encouraged to make graph reading a "meta-cognitive" activity, that is, to pay attention to and reflect critically on the interpretation of graphs.

One of the typical problems found in graph comprehension, even in higher education, is the interpretation of graphs as "concrete iconic" objects, that is, images rather than relational structures. An example found in students of chemical engineering at DTU were errors in reasoning about a test on Fourier's law for heat conduction, where some students, who tried to remember the graph by its shape (a graphical image feature) rather than by the conceptual model it expresses, reported the correct shape of a curved graph, but as a mirror image of its correct orientation (May, 1998).

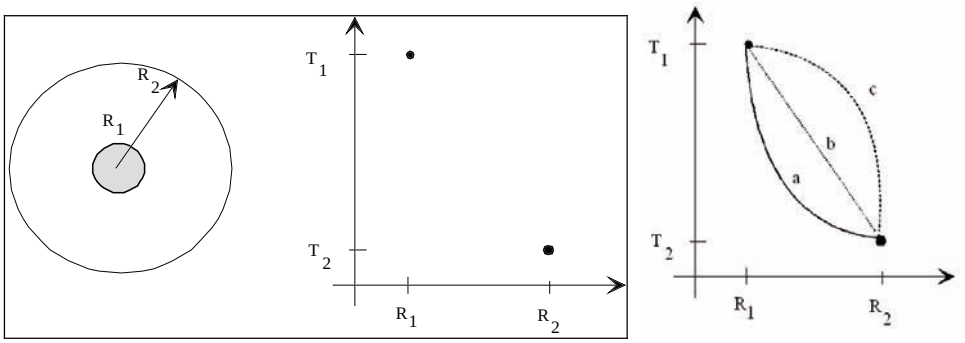
The students who answered (c) to the test in Figure 19 could not give any physical interpretation to the graph since it was remembered as a mere image of a textbook (or lecture) presentation. This exemplifies a "typological error" within the representational forms used in reasoning by the students: treating a graph as if it was an image (or effectively reducing it to a mental image).

In order to support the didactic suggestions by Shah and Hoeffner (2002) and others regarding the need to train graph comprehension skills in higher education, learning objects need to be annotated with metadata on their level of interactivity (to distinguish learning objects with active exploration of models from more or less static illustrations, for example) and on their support for active student learning through translations between different representational forms (as in conceptual tests motivating students to transform graph representations, algebraic expressions, or natural language descriptions of models).

The taxonomy of representational forms introduced in this chapter can be used to extend existing standards for learning object metadata with this type of information (as well as specifications of learning objectives).

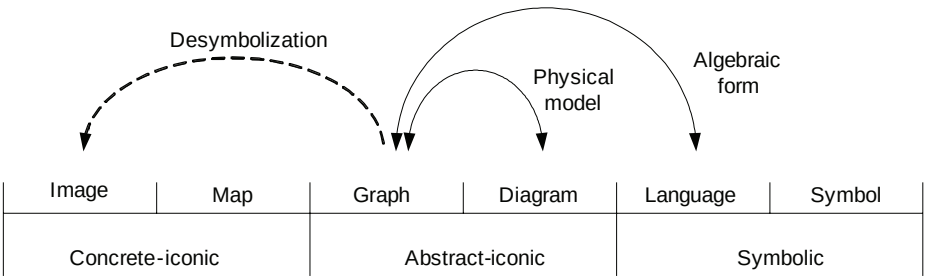
In a recent study, it was found that students working actively with integrating multiple static representations, before working with animations and simulations, have better learning outcomes than students who directly explore dynamic and interactive visualisations (Bodemer, Ploetzner, Bruchmüller, & Häcker, 2005). This is again empirical evidence that indicates the importance of the integration of multiple forms of representation for conceptual understanding. In the context of instructional design of learning objects for object repositories, it will be necessary to extend current XML-based standards for metadata with infor-

Figure 19. A conceptual test (to the left + caption) given to students in a course in chemical engineering at DTU (1995); three types of answers were given by students: the correct answer (a), the linear answer (b), which is motivated by the original context where students learned about the law (where it is in fact linear, if the area is not considered), and (c) where student have remembered the shape of the image, but without any physical understanding of the model represented by the graph (May, 1998)



[Original caption to the test:] Consider a cylinder with radius R_2. Inside the cylinder is a kernel with radius R_1 and which has been heated to the temperature T_1. At the surface (with radius $r = R_2$) the temperture is held constant at T_2 where $T_2 < T_1$. Show in the diagram a qualitative sketch of the graph T(r) for values between R_1 and R_2 (i.e., you do not have to compute any values for points on the curves).

Figure 20. Conceptual diagram to indicate the "desymbolisation" of graphs reduced to concrete-iconic images by some students and the didactic importance of training students actively in "translating" between graphs and their "diagrammatic" interpretation in terms of their underlying physical models as well as between graphs and the formal algebraic models which they express



mation about the representational forms utilised by learning objects. This will enable teachers engaged in instructional design of learning objects to make better choices in setting up conditions for active learning, since students will benefit from being exposed to many different representations of the same content. In engineering education, for instance, it is important to set up learning environments where students can work actively with static graphs, algebraic forms, explanations in natural language, and conceptual diagrams, as well as animated graphs and interactive simulations: Conceptual understanding emerges from the discovery of invariant forms across all of these media and "modalities."

SUMMARY

A better understanding of the invariant semantic properties of representational forms across different media will be important in the future support for flexible configurable interfaces and for supporting the design of device-independent multimedia and other forms of adaptive multimedia.

In the domain of instructional design and in the redesign of instructional material as learning objects for Web-based learning and for learning repositories, it is of primary importance to extend current standards for learning objects with (a) didactic specifications of the intended use of these objects and (b) semiotic and semantic information about the media and representational forms supported by the learning objects.

An approach to the taxonomy of media types and representational forms has been suggested, based on feature structures and Formal Concept Analysis. More research is needed in order to combine the analysis of media types and representational forms with the analysis of scale types and scale transformations (Petersen & May, in press), and to apply both to the domain of adaptive and configurable interfaces as well as to the domain of instructional design.

REFERENCES

Anderson, T., & Petrinjak, A. (2004). Beyond learning objects to educational modelling languages. In R. McGreal (Ed.), *Online education using learning objects* (pp. 287-300). London; New York: RoutledgeFalmer.

Bateman, J., Delin, J., & Henschel, R. (2002). *XML and multimodal corpus design: Experiences with multi-layered stand-off annotations in the GeM corpus.* ELSNET Workshop (Towards a Roadmap for Multimodal Language Resources and Evaluation) at the LREC 2002 Conference, Las Palmas, Spain.

Bateman, J., Kamps, T., Kleinz, J., & Reichenberger, K. (2001). Towards constructive text, diagram, and layout generation for information presentation. *Computational Linguistics, 27*(3), 409-449.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics, 62*(8), 750-762

Bernsen, N. O. (1993, November). *Modality theory: Supporting multimodal interface design.* ERCIM Workshop on Multimodal Human-Computer Interaction (Report ERCIM-94-W003), Nancy, France.

Bernsen, N. O. (1994). Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers, 6*(4), 347-371.

Bodemer, D., Ploetzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science, 33*, 73-95.

Bulterman, D. (2001). SMIL 2.0. Part 1: Overview, concepts, and structure. *IEEE Multimedia, 8*(4), 82-88.

Chatman, S. (1978). *Story and discourse: Narrative structure in fiction and film*. Ithaca, New York: Cornell University Press.

Cox, R. (1999). Representation construction, externalised cognition, and individual differences. *Learning and Instruction, 9*, 343-363.

Davey, B. A., & Priestley, H. A. (1990). *Introduction to lattices and order.* Cambridge, UK: Cambridge University Press.

de Saussure, F. (1993). *Course in general linguistics.* London: Dockworth.

De Westelinck, K., Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in the social sciences: Limitations of external graphical

representations. *Computers in Human Behavior, 21*, 555-573.

Dorai, C., & Venkatesh, S. (2001, September 10-12). Bridging the semantic gap in content management systems: Computational media aesthetics. *Proceedings of the First Conference on Computational Semiotics, COSIGN-2001*, Amsterdam. Retrieved from http://www.cosign-conference.org/cosign2001

Dorai, C., & Venkatesh, S. (2003). Bridging the semantic gap with computational media aesthetics. *IEEE Multimedia, 10*(2), 15-17.

Friesen, N. (2004). Three objections to learning objects. In R. McGreal (Ed.), *Online education using learning objects* (pp. 59-70). Routledge-Falmer.

Ganter, B., & Wille, R. (1999). *Formal concept analysis: Mathematical foundations*. Berlin: Springer.

Gourdol, A., Nigay, L., Salber, D., & Coutaz, J. (1992). Two case studies of software architecture for multimodal interactive systems. *Proceedings of the IFIP Working Conference on Engineering for Human Computer Interaction* (pp. 271-284). Ellivuori, Finland: Amsterdam: North-Holland.

Heller, R. S., & Martin, C. D. (1995). A media taxonomy. *IEEE Multimedia, 2*(4), 36-45.

Heller, R. S., Martin, C. D., Haneef, N., & Gievska-Krliu, S. (2001). Using a theoretical multimedia taxonomy framework. *ACM Journal of Educational Resources in Computing, 1*(1), 1-22.

Hende, A. (2002). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia, 11*(2), 177-191.

Hjelmslev, L. (1961). *Omkring Sprogteories Grundlæggelse. Travaux du Cercle Linguistique du Copenhague, XXV (Reissued 1993, Reitzel).* English translation in L. Hjelmslev, *Prolegomena to a theory of language*. Bloomington: Indiana University Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Koper, R. (2001). *Modelling units of study from a pedagogical perspective: The pedagogical meta-model behind EML*. The Educational Technology Expertise Centre of the Open University of the Netherlands. Retrieved from http://eml.ou.nl/introduction/docs/ped-metamodel.pdf

Koper, R., & van Es, R. (2004). Modelling units of study from a pedagogical perspective. In R. McGreal (Ed.), *Online education using learning objects* (pp. 43-58). London; New York: RoutledgeFalmer.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University Of Chicago Press.

Lakoff, G. (1990). The invariance hypothesis: Is abstract reasoning based on image-schemas? *Cognitive Linguistics, 1*(1), 39-74.

Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. (1994). A classification of visual representations. *Communications of the ACM, 37*(12), 36-49.

Lohse, G. L., Walker, N., Biolsi, K., & Rueter, H. (1991). Classifying graphical information. *Behaviour & Information Technology, 10*(5), 419-436.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics, 5*(2), 110-141.

Mackinlay, J., & Genesereth, M. R. (1985). Expressiveness and language choice. *Data & Knowledge Engineering, 1*(1), 17-29.

Manning, A. D. (1989). The semantics of technical graphics. *Journal of Technical Writing and Communication, 19*(1), 31-51.

May, M. (1993). A taxonomy of representations for HCI, Parts 1-3. In N. O. Bernsen (Ed.), *Taxonomy of HCI systems: State of the art.* Esprit

basic research project GRACE working papers. Deliverable 2.1. Edinburgh: Human Communication Research Centre.

May, M. (1998). Images, diagrams, and metaphors in science and science education: The case of chemistry. *Almen Semiotik, 14,* 77-102.

May, M. (1999). Diagrammatic reasoning and levels of schematization. In T. D. Johansson, M. Skov, & B. Brogaard (Eds.), *Iconicity: A fundamental problem in semiotics.* Copenhagen: NSU Press.

May, M. (2001, June 25-27). Semantics for instrument semiotics. In M. Lind (Ed.), *Proceedings of the 20th European Annual Conference on Human Decision Making and Manual Control (EAM-2001),* Kongens Lyngby, Denmark (pp. 29-38).

May, M. (2001). Instrument semiotics: A semiotic approach to interface components. *Knowledge-Based Systems, 14*(2001), 431-435.

May, M. (2004). Wayfinding, ships, and augmented reality. In P. B. Andersen & L. Qvortrup (Eds.), *Virtual applications: Applications with virtual inhabited 3D worlds.* Berlin; New York: Springer Verlag.

May, M., & Andersen, P. B. (2001). Instrument semiotics. In K. Liu, R. J. Clarke, P. B. Andersen, & R. K. Stamper (Eds.), *Information, organisation, and technology: Studies in organisational semiotics.* Boston; Dordrecht; London: Kluwer Academic Publishers.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge, UK: Cambridge University Press.

McDermott, L. C., Rosenquist, M. L., & van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics, 55*(6), 503-513.

Nack, F., & Hardman, L. (2002). *Towards a syntax for multimedia semantics* (Rep. No. INS-RO204). Amsterdam: Information Systems (INS) at CWI.

Nigay, L., & Coutaz, J. (1993). A design space for multimodal systems: Concurrent processing and data fusion. *INTERCHI'93 Proceedings.* Amsterdam: ACM Press; Addison Wesley.

Peirce, C. S. (1906). Prolegomena for an apology for pragmatism. In C. Eisele (Ed.), *The new elements of mathematics, Vol. IV.* The Hague: Mouton.

Petersen, J., & May, M. (in press). Scale transformations and information presentation in supervisory control. *International Journal of Human-Machine Studies.* Retrieved from www.sciencedirect.com

Purchase, H. C., & Naumann, D. (2001). A semiotic model of multimedia: Theory and evaluation. In S. M. Rahman (Ed.), *Design and management of multimedia information systems: Opportunities and challenges* (pp. 1-21). Hershey, PA: Idea Group Publishing.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X. -F., Kirbas, C., McCullough, K. E., & Ansari, R. (2002). Multimodal human discourse: Gesture and speech. *ACM Transactions on Computer-Human Interaction, 9*(3), 171-193.

Recker, M. M., Ram, A., Shikano, T., Li, G., & Stasko, J. (1995). Cognitive media types for multimedia information access. *Journal of Educational Multimedia and Hypermedia, 4*(2-3), 183-210.

Schnotz, W., & Bannert, M. (2003): Construction and interference in learning from multiple representations. *Learning and Instruction, 13,* 141-156.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review, 14*(1), 47-69

Stenning, K., Inder, R., & Neilson, I. (1995). Applying semantic concepts to analyzing media and modalities. In J. Glasgow, N. H. Narayanan, & B. Chandrasekaran (Eds.), *Diagrammatic reason-*

ing: Cognitive and computational perspectives. Menlo Park, CA: AAAI Press; MIT Press.

Stockinger, P. (1993). Multimedia and knowledge based systems. *S – European Journal of Semiotic Studies, 5*(3), 387-424.

Troncy, R., & Carrive, J. (2004). A reduced yet extensible audio-visual description language. *ACM Symposium on Document Engineering* (pp. 87-89), Milwaukee, Wisconsin.

Twyman, M. (1979). A schema for the study of graphic language. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language, Vol. 1* (pp. 117-150). New York: Plenum Press.

Ware, C. (2000). *Information visualization: Perception for design.* San Francisco: Morgan Kaufman.

Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), *The instructional use of learning objects.* Retrieved from http://reusability.org/read/

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human-Computer Studies, 45,* 59-74.

This work was previously published in Digital Multimedia Perception and Design, edited by G. Ghinea and S.Y. Chen, pp. 47-80, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.8
Cognitively Informed Multimedia Interface Design

Eshaa M. Alkhalifa
University of Bahrain, Bahrain

INTRODUCTION

The rich contributions made in the field of human computer interaction (HCI) have played a pivotal role in shifting the attention of the industry to the interaction between users and computers (Myers, 1998). However, technologies that include hypertext, multimedia, and manipulation of graphical objects were designed and presented to the users without referring to critical findings made in the field of cognitive psychology. These findings allow designers of multimedia educational systems to present knowledge in a fashion that would optimize learning.

BACKGROUND

The long history of human computer interaction (HCI) has witnessed many successes represented in insightful research that finds its way to users' desktops. The field influences the means through which users interact with computers—from the introduction of the mouse (English et al., 1967) and applications for text editing (Meyrowitz & Van Dam, 1982) to comparatively recent areas of research involving multimedia systems (Yahaya & Sharifuddin, 2000).

Learning is an activity that requires different degrees of cognitive processing. HCI research recognized the existence of diversity in learning

styles (Holt & Solomon, 1996) and devoted much time and effort toward this goal. However, Ayre and Nafalski (2000) report that the term *learning styles* is not always interpreted the same way and were able to offer two major interpretations. The first group believes that learning styles emerge from personality differences, life experiences, and student learning goals, while the second group believes that it refers to the way students shape their learning method to accommodate teacher expectations, as when they follow rote learning when teachers expect it.

The first interpretation includes, in part, a form of individual differences but does not explicitly link them to individual cognitive differences, which, in turn, caused researchers more ambiguities as to interpreting the different types of learning styles. In fact, these differences in interpretations caused Stahl (1999) to publish a critique, where he cites five review papers that unite in concluding the lack of sufficient evident to support the claim that accommodating learning styles helps to improve children's learning when acquiring the skill to read. He criticized Carbo's reading style inventory and Dunn and Dunn's learning inventory because of their reliance on self-report to identify different learning styles of students, which, in turn, results in very low replication reliability.

These criticisms are positive in that they indicate a requirement to base definitions on formal replicable theory. A candidate for this is cognitive learning theory (CLT), which represents the part of cognitive science that focuses on the study of how people learn the information presented to them and how they internally represent the concepts mentally in addition to the cognitive load that is endured during the learning process of the concepts.

Some of the attempts that were made to take advantage of the knowledge gained in the field include Jonassen (1991), van Jooligan (1999), and Ghaoui and Janvier (2004).

Jonassen (1991) advocates the constructivist approach to learning, where students are given several tools to help them perform their computation or externally represent text they are expected to remember. This allows them to focus on the learning task at hand. Jonassen (1991) adopts the assumption originally proposed by Lajoie and Derry (1993) and Derry (1990) that computers fill the role of cognitive extensions by performing tasks to support basic thinking requirements, such as calculating or holding text in memory, and thus allowed computers to be labeled cognitive tools. Jonassen's (1991) central claim is that these tools are offered to students to lower the cognitive load imposed during the learning process, which, in turn, allows them to learn by experimentation and discovery.

Van Jooligan (1999) takes this concept a step further by proposing an environment that allows students to hypothesize and to pursue the consequences of their hypotheses. He did this through utilizing several windows in the same educational system. The system was composed of two main modules: the first supports the hypothesis formation step by providing menus to guide the process; the second provides a formatted presentation of experiments already tested and their results in a structured manner. They also added intelligent support to the system by providing feedback to students to guide their hypothesis formation approach.

Ghaoui and Janvier (2004) presented a two-part system. The first part identified the various personality types, while the second either had an interactive or non-interactive interface. They report an increase in memory retention from 63.57% to 71.09% that occurred for the students using the interactive interface. They also provided a description of the learning style preferences for the students tested, which exhibited particular trends, but these were not analyzed in detail.

Montgomery (1995) published preliminary results of a study aimed at identifying how mul-

timedia, in particular, can be used to address the needs of various learning styles. Results indicate that active learners appreciate the use of movies and interaction, while sensors benefit from the demonstrations.

Although a glimmer of interest in CLT exists, there is a distinct lack of a clear and organized framework to help guide educational interface designers.

ALIGNMENT MAP FOR MULTIMEDIA INSTRUCTIONAL INTERFACE

The problems that arose with learning styles reveal a need for a more fine-grained isolation of various cognitive areas that may influence learning. Consequently, an alignment map, as shown in Table 1, may offer some guidelines as to what aspects of the multimedia interface design would benefit from what branch of the theory in order to gain a clearer channel of communication between the designer and the student.

CASE STUDY: DATA STRUCTURES MULTIMEDIA TUTORING SYSTEM

The alignment map presents itself as an excellent basis against which basic design issues of multimedia systems may be considered with the goal of making the best possible decisions.

The multimedia tutoring system considered here (Albalooshi & Alkhalifa, 2002) teaches data structures and was designed by considering the various design issues as dictated by the alignment map that was specifically designed for the project and is shown in Table 1. An analysis of the key points follows:

1. **Amount of Media Offered:** The system presents information through textual and animated presentation only. This is done to avoid cognitive overload caused by redundancy (Jonassen, 1991) that would cause students to find the material more difficult to comprehend.
2. **How the Screen is Partitioned:** The screen grants two-thirds of the width to the animation window that is to the left of the screen, while the verbal description is to the right. Although the language used for the textual description is in English, all students are Arabs, so they are accustomed to finding the text on the right side of the screen, because in Arabic, one starts to write from the right hand side. This design, therefore, targeted this particular pool of students to ensure that both parts of the screen are awarded sufficient attention. It presents an interface that requires divided attention to two screens that complement each other, a factor that, according to Hampson (1989), minimizes interference between the two modes of presentation.
3. **Parallel Delivery of Information:** Redundancy is desired when it exists in two different media, because one re-enforces the other. It is not desired when it exists within the media, as when there is textual redundancy and things are explained more than once. Consequently, the textual description describes what is presented in the animation part, especially since only text and animation media exist in this case, which means that cognitive load issues are not of immediate concern (Jonassen, 1991).
4. **Use of Colors:** Colors were used to highlight the edges of the shapes and not on a wide scale to ensure that attention is drawn to those. By doing so, focus is expected to be directed toward the object's axes, as suggested by Marr and Nishihara (1978), in order to encourage memory recall of the shapes at a later point in time.
5. **Use of Animation:** The animated data structures are under the user's control with

Table 1. Alignment map from multi-media design questions to various cognitive research areas that may be of relevance

Multimedia Design Issues	Cognitive Areas That May Be of Relevance
1. Amount of media offered 1	. Cognitive load 2. Limited attention span 3. Interference between different mental representations
2. How the screen is partitioned 1	. Perception and recognition 2. Attention
3. Parallel delivery of information	1. Redundancy could cause interference 2. Limited working memory (cognitive load issues) 3. Limited attention span 4. Learner difference
4. Use of colors	1. Affects attention focus 2. Perception of edges to promote recall
5. Use of animation	1. Cognitive load reduction 2. Accommodates visualizer/verbalizer learners
6. Use of interactivity	1. Cognitive load reduction 2. Raises the level of learning objectives
7. Aural media	1. Speech perception issues like accent and clarity 2. Interference with other media
8. Verbal presentation of material	1. Clarity of communication 2. Accommodates verbal/serialist learners

respect to starting, stopping, or speed of movement. This allows the user to select whether to focus on the animation, text, or both in parallel without causing cognitive overload.

6. **Use of Interactivity:** The level of interactivity is limited to the basic controls of the animation.
7. **Aural Media:** This type of media is not offered by the system.
8. **Verbal Presentation of Material:** The verbal presentation of the materials is concise and fully explains relevant concepts to a sufficient level of detail, if considered in isolation of the animation.

EVALUATION OF THE SYSTEM

The tool first was evaluated for its educational impact on students. It was tested on three groups: one exposed to the lecture alone; the second to a regular classroom lecture in addition to the system; and the third only to the system. Students were distributed among the three groups such that each group had 15 students with a mean grade similar to the other two groups in order to ensure that any learning that occurs is a result of the influence of what they are exposed to. This also made it possible for 30 students to attend the same lecture session composed of the students of groups one and two, while 30 students attended the same lab session composed of the students of groups two and three in order to avoid any confounding factors.

Results showed a highly significant improvement in test results of the second group when their post-classroom levels were compared to their levels following use of the multimedia system, with an overall improvement rate of 40% recorded with $F=9.19$, with $p < 0.005$ from results of an ANOVA test after ensuring all test requirements had been

satisfied. The first and third groups showed no significant differences between them.

Results shown indicate that learning did occur to the group that attended the lecture and then used the system, which implies that animation does fortify learning by reducing the cognitive load. This is especially clear when one takes the overall mean grade of all groups, which is around 10.5, and checks how many in each group are above that mean. Only six in group one were above it, while 11 were above it in group two and 10 in group three. Since group three was exposed to the system-only option and achieved a number very close to group two, which had the lecture and the system option, then clearly, multimedia did positively affect their learning rate.

Since one of the goals of the system is to accommodate learner differences, a test was run on group two students in order to identify the visualizers from the verbalizers. The paper-folding test designed by French et al. (1963) was used to distinguish between the two groups. The test requires each subject to visualize the array of holes that results from a simple process. A paper is folded a certain number of folds, a hole is made through the folds, and then the paper is unfolded. Students are asked to select the image of the unfolded paper that shows the resulting arrangement, and results are evaluated along a median split as high vs. low visualization abilities.

These results then were compared with respect to the percentage of improvement, as shown in Table 2. Notice that the question numbers in the pre-test are mapped to different question numbers in the post-test in order to minimize the possibility of students being able to recall them; a two-part question also was broken up for the same reason.

Results indicate that, although the group indeed was composed of students with different learning preferences, they all achieved comparable overall improvements in learning. Notice, though, the difference in percentage improvement in Question 4. The question is: List and explain the data variables that are associated with the stack and needed to operate on it. This particular question is clearly closer to heart to the verbalizer group than to the visualizer group. Therefore, it should not be surprising that the verbalizer group finds it much easier to learn how to describe the data variables than it is for students who like to see the stack in operation. Another point to consider is that the visualizer group made a bigger improvement in the Q1+Q8 group in response to the question: Using an example, explain the stack concept and its possible use. Clearly, this question is better suited to a visualizer than to a verbalizer.

FUTURE TRENDS

CLT already has presented us with ample evidence of its ability to support the design of more informed

Table 2. The percentage improvement of each group from the pretest to the posttest across the different question types

	Q1 PLUS Q8 MAPPED TO Q1	Q3 MAPPED TO Q2	Q4 MAPPED TO Q3	Q6 MAPPED TO Q6
VISUALIZER GROUP	27.8%	18.6%	9.72%	9.76%
T-TEST RESULTS	.004	.003	.09	.01
VERBALIZER GROUP	20.7%	22.8%	21.4%	15.7%
T-TEST RESULTS	.004	.005	.003	.009

and, therefore, more effective educational systems. This article offers a map that can guide the design process of a multimedia educational system by highlighting the areas of CLT that may influence design. The aim, therefore, is to attract attention to the vast pool of knowledge that exists in CLT that could benefit multimedia interface design.

CONCLUSION

This article offers a precise definition of what is implied by a computer-based cognitive tool (CT) as opposed to others that were restricted to a brief definition of the concept. Here, the main features of multimedia were mapped onto cognitive areas that may have influence on learning, and the results of an educational system that conforms to these design requirements were exhibited.

These results are informative to cognitive scientists, because they show that the practical version must deliver what the theoretical version promises. At the same time, results are informative to educational multimedia designers by exhibiting that there is a replicated theoretical groundwork that awaits their contributions to bring them to the world of reality.

The main conclusion is that this is a perspective that allows designers to regard their task from the perspective of the cognitive systems they wish to learn so that it shifts the focus from a purely teacher-centered approach to a learner-centered approach without following the route to constructivist learning approaches.

REFERENCES

AlBalooshi, F., & Alkhalifa, E. M. (2002). Multi-modality as a cognitive tool. *Journal of International Forum of Educational Technology and Society, IEEE, 5*(4), 49-55.

Ayre, M., & Nafalski, A. (2000). Recognising diverse learning styles in teaching and assessment of electronic engineering. *Proceedings of the 30th ASEE/IEEE Frontiers in Education Conference*, Kansas City, Missouri.

Derry, S. J. (1990). Flexible cognitive tools for problem solving instruction. *Proceedings of the Annual Meeting of the American Educational Research Association*, Boston.

English, W. K., Engelbart, D. C., & Berman, M. L. (1967). Display selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics, 8*(1), 5-15.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Services.

Ghaoui, C., & Janvier, W. A. (2004). Interactive e-learning. *Journal of Distance Education Technologies, 2*(3), 23-35.

Hampson, P. J. (1989). Aspects of attention and cognitive science. *The Irish Journal of Psychology, 10*, 261-275.

Jonassen, D. H. (1991). Objectivism vs. constructivism: Do we need a new philosophical paradigm shift? *Educational Technology: Research and Development, 39*(3), 5-13.

LaJoie, S. P., & Derry, S. J. (Eds.). (1993). *Computers as cognitive tools*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marr, D., & Nishihara, K. (1978). *Representation and recognition of the spatial organization of three-dimensional shapes*. London: Philosophical Transactions of the Royal Society.

Meyrowitz, N., & Van Dam, A. (1982). Interactive editing systems: Part 1 and 2. *ACM Computing Surveys, 14*(3), 321-352.

Montgomery, S. M. (1995). Addressing diverse learning styles through the use of multimedia.

Proceedings of the 25th ASEE/IEEE Frontiers in Education Conference, Atlanta, Georgia.

Myers, B. A. (1998). A brief history of human computer interaction technology. *ACM Interactions, 5*(2), 44-54.

Stahl, S. (1999). Different strokes for different folks? A critique of learning styles. *American Educator, 23*(3), 27-31.

Yahaya, N., & Sharifuddin, R. S. (2000). Concept-building through hierarchical cognition in a Web-based interactive multimedia learning system: Fundamentals of electric circuits. *Proceedings of the Ausweb2k Sixth Australian World Wide Web Conference*, Cairns, Australia.

KEY TERMS

Alignment Map: A representation on a surface to clearly show the arrangement or positioning of relative items on a straight line or a group of parallel lines.

Attention: An internal cognitive process by which one actively selects which part of the environmental information that surrounds them and focuses on that part or maintains interest while ignoring distractions.

Cognitive Learning Theory: The branch of cognitive science that is concerned with cognition and includes parts of cognitive psychology, linguistics, computer science, cognitive neuroscience, and philosophy of mind.

Cognitive Load: The degree of cognitive processes required to accomplish a specific task.

Learner Differences: The differences that exist in the manner in which an individual acquires information.

Multimedia System: Any system that presents information through different media that may include text, sound, video computer graphics, and animation.

This work was previously published in Encyclopedia of Human Computer Interaction, edited byC. Ghaoui, pp. 79-84, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.9
Enabling Multimedia Applications in Memory-Limited Mobile Devices

Raul Fernandes Herbster
Federal University of Campina Grande, Brazil

Hyggo Almeida
Federal University of Campina Grande, Brazil

Angelo Perkusich
Federal University of Campina Grande, Brazil

Marcos Morais
Federal University of Campina Grande, Brazil

INTRODUCTION

Embedded systems have several constraints which make the development of applications for such platforms a difficult task: memory, cost, power consumption, user interface, and much more. These characteristics restrict the variety of applications that can be developed for embedded systems. For example, storing and playing large videos with good resolution in a limited memory and processing power mobile device is not viable.

Usually, a client-server application is developed to share tasks: clients show results while servers process data. In such a context, another hard task for limited memory/processing devices could be delegated to the server: storage of large data. If the client needs data, it can be sent piece by piece from the server to the client.

In this article we propose a layered architecture that makes possible the visualization of large videos, and even other multimedia documents, in memory/processing limited devices. Storage of videos is performed at the server side, and the client plays the video without worrying about

storage space in the device. Data available in the server is divided into small pieces of readable data for mobile devices, generally JPEG files. For example, when the client requests videos from the server, the videos are sent as JPEG files and shown at an ideal rate for users. The video frames are sent through a wireless connection.

The remainder of this article is organized as follows. We begin by describing background concepts on embedded systems and client-server applications, and then present our solution to enable multimedia applications in memory-limited mobile devices. We next discuss some future trends in mobile multimedia systems, and finally, present concluding remarks.

BACKGROUND

Embedded Systems

An embedded system is not intended to be a general-purpose computer. It is a device designed to perform specific tasks, including a programmable computer. A considerable number of products use embedded systems in their design: automobiles, personal digital assistants, and even household appliances (Wayne, 2005). These limited systems have some constraints that must be carefully analyzed while designing the applications for them: size, time constraints, power consumption, memory usage and disposal, and much more (Yaghmour, 2003).

These constraints restrict the variety of software for embedded systems. The development of applications which demand a large amount of memory, for example, is not viable for embedded systems, because the memory of such devices is limited. Extra memory can also be provided, but the total cost of application is very high. Another example is multimedia applications, such as video players: storing and playing large videos with good resolution in a limited memory and processing power mobile device is a very hard task.

There are specific platforms that were developed to perform multimedia tasks: embedded video decoders and embedded digital cameras, for example. However, other considerable parts of embedded systems, like personal digital assistants (PDAs) and cell phones, are not designed to play videos with good quality, store large amount of data, and encode/decode videos. Thus, it is important to design solutions enabling multimedia environments in this variety of memory/processing-limited devices.

Layered and Client-Server Architectures

Layered architectures share services through a hierarchical organization: each layer provides specific services to the layers above it and also acts as a client to the layer below (Shaw & Garlan, 1996). This characteristic increases the level of abstraction, allowing the partition of complex problems into a set of tasks easier to perform. Layered architectures also decouple modules of the software, so reuse is also more easily supported. As communication of layers is made through contracts specified as interfaces, the implementation of each module can be modified interchangeably (Bass & Kazman, 1998).

Most of the applications have three major layers with different functionalities: presentation, which handles inputs from devices and outputs to screen display; application or business logic, which has the main functionalities of the application; and data, which provides services for storing the data of the application (Fastie, 1999).

The client-server architecture has two elements that establish communication with each other: the front-end or client portion, which makes a service request to another program, called server; and the back-end or server portion, which provides service to the request. The client-server architecture allows an efficient way to interconnect programs that are distributed at different places (Jorwekar, 2005). However, the client-server architecture

is more than just a separation of a user from a server computer (Fastie, 1999). Each portion has also its own modules: presentation, application, and data.

ENABLING MULTIMEDIA APPLICATIONS

Multimedia applications demand a considerable amount of resources from the environment in order to guarantee quality of service, which can be defined in terms of security, availability, or efficiency (Banâtre, 2001). Embedded systems have several constraints, like limited memory (Yaghmour, 2003), which make it very difficult to implement multimedia applications in an embedded platform.

Today, the growing interest for mobile devices and multimedia products requires the development of multimedia applications for embedded systems (Banâtre, 2001). There are approaches (Grun, Balasa, & Dutt, 1998; Leeman et al., 2005) that try to enhance embedded systems memory and other system aspects, such as processing, to provide better results in multimedia applications. However, most of the solutions available focus on hardware architecture, and a large number of programmers are not used to programming at the hardware level.

A solution based on client-server architecture is a good proposal for limited-memory/processing mobile devices because harder tasks can be performed by the server side whereas the client just displays results. By designing applications based on an architecture that shares tasks, constraints like limited memory and low computing power are partially solved. In this article, we propose a layered, client/server architecture that allows playing and storing large videos on limited-memory/processing mobile devices. The data is sent through a wireless intranet.

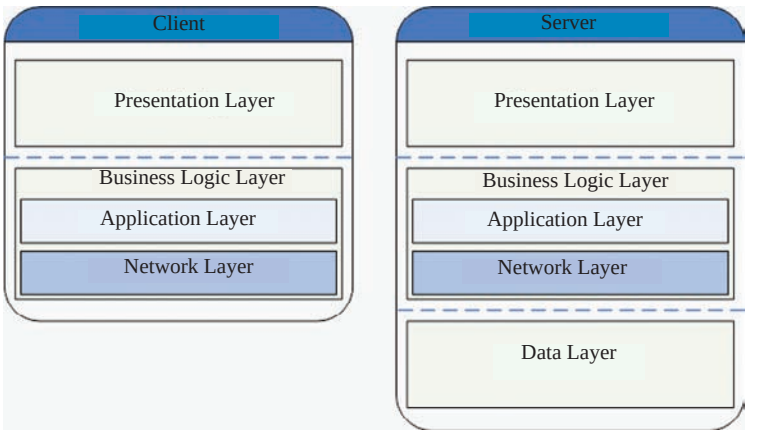
Client/Server Architecture

In Figure 1, both client and server modules are illustrated. Each module of both elements can be changed at any time, except the application layer because the rules of application are defined on it: if business logic changes, so does the application.

The server architecture is a standard three-tier architecture:

- **Presentation Layer:** This layer interacts directly with the client. Its functionalities are related to display forms so that the user adds multimedia content to the server repository.

Figure 1. Client/server architecture

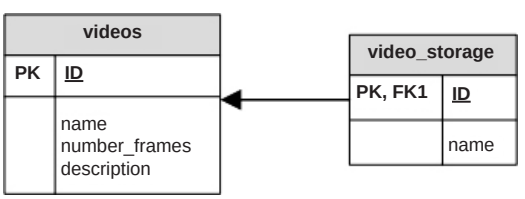


- **Business Logic Layer:** This has two sublayers: the network layer, which manages the connection of server and client, receiving requests and sending responses; and the application layer, which requests services for the data layer, such as document storage and reports.
- **Data Layer:** This layer stores the multimedia documents as JPEG files. Each document has information about management files, including ID, number of frames (JPEG files), and specific description elements, which depends on the multimedia document type. Figure 2 illustrates the logic model of the tables that contain such information. The server stores in a table (video_storage) the titles and ID of all videos. All the information about a video is stored in another table (videos).

The client is a single two-tier application; the data layer is not defined.

- **Presentation Layer:** This consists of a video player with buttons to select the video and a screen to display the video.
- **Business Logic Layer:** This has two sublayers: the network layer, which manages the connection with the client, sending requests to the server and receiving data from it; and the application layer, which gets frames from the network layer and also controls the tax rate of displaying the frames.

Figure 2. Logic model of video descriptions information



Execution Scenario

In what follows, we present an execution scenario of the mobile multimedia architecture. For this, consider that, at server side, a video was divided into small pieces of readable data (JPEG files). The quantity of pieces depends on the desired quality of the video that will be played at the client side.

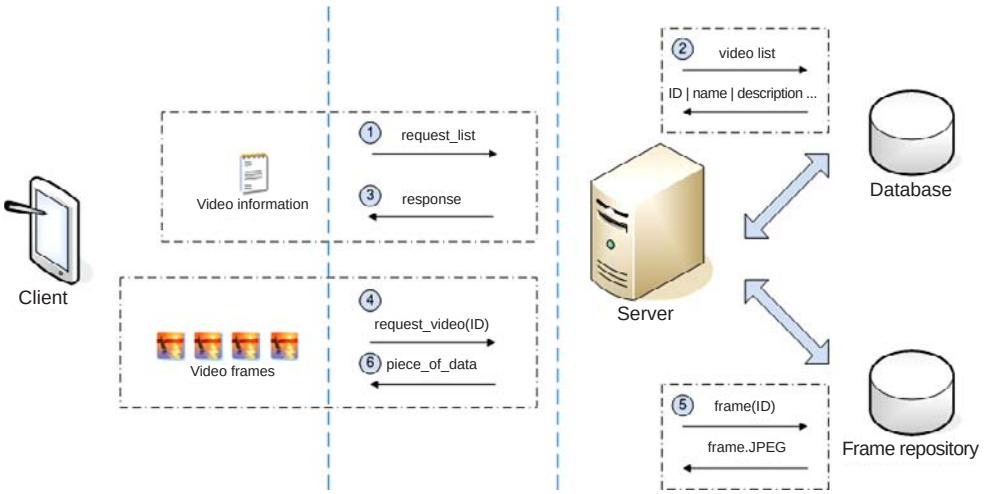
The communication process between client and server is illustrated in Figure 3 and is described as follows. Whenever the server receives a connection request from a client, it sends to the client the list of available videos (steps 1 and 2). The client receives data and displays this information whenever required (step 3). Then, the client sends to the server the ID of the requested video (step 4). The server receives the request and starts the transmission of the video, piece by piece (steps 5 and 6).

At the client side, the pieces of data (JPEG files) are received in a tax rate that depends on the network (traffic, band, etc.). However, to display frames to the client in a constant tax rate and guarantee quality of service, it is necessary to maintain a buffer, which is controlled by the video player.

In the architecture described, the data is sent through a wireless intranet. The pieces of data are JPEG files, but could also be in Motion JPEG (MJPEG) format. The tax of frames depends on the quality of the video player on the client side: generally, a high video quality rate is 30 frames per second, whereas a low video quality rate is 10 frames per second.

In a wireless network, this solution needs a large part of the network bandwidth. Therefore, there is a tradeoff between memory/processing capacity and network bandwidth. Nevertheless, considering home entertainment environments, such a tradeoff is worth the cost mainly because wireless networks in home environments have enough bandwidth to be used in such a context.

Figure 3. Client/server communication



The architecture can also be used for other kinds of multimedia documents. For example, large PDF documents cannot be visualized with good quality on memory/processing-limited devices. The PDF documents can be also shared as JPEG files and sent to the client.

FUTURE TRENDS

A protocol defines communication between client and server. It is an important element to guarantee QoS. As for future work, we suggest a deeper study of protocols enabling a good service for a given situation. It is important to focus on protocols that do not demand a lot of resources from the network, such as bandwidth.

There are some protocols that are implemented over UDP, for example, trivial file transport protocol (TFTP) (RFC 783, 1981) and real-time transfer protocol (RTP) (RFC, 1996). These protocols were implemented to demand few resources of the network, and to transfer a considerable number of files through the network.

Another interesting research approach is to measure variables of the network while using an application based on the architecture described, for example, to define how many devices running such applications the network supports.

CONCLUSION

Multimedia applications demand a considerable amount of resources from systems. To develop multimedia applications for embedded systems, it is necessary to tackle constraints inherent to such platforms, such as limited memory and processing power.

In this article, we described a general architecture used for enabling multimedia applications in memory/processing systems. The architecture has two parts: a server, which receives requests and sends responses to clients; and clients, which make requests. Both parts have a layered architecture.

The solution proposed is relatively simple to implement and is easy to maintain, because the modules are decoupled and can be modified interchangeably. However, the architecture demands a considerable amount of network bandwidth, because the number of packages sent by the server to the client is large. Nevertheless, considering that the application is implemented over a wireless

network in home environments, the bandwidth tradeoff is worth the cost.

REFERENCES

Banâtre, M. (2001). Ubiquitous computing and embedded operating systems design. *ERCIM News*, (47).

Bass, C., & Kazman. (1998). *Software architecture in practice*. Boston: Addison Wesley Longman.

Fastie, W. (1999). Understanding client/server computing. *PC Magazine*, 229-230.

Grun, P., Balasa, F., & Dutt, N. (1998). Memory size estimation for multimedia applications. *International Conference on Hardware Software Codesign, Proceedings of the 6th International Workshop on Hardware/Software Codesign* (pp. 145-149), Seattle, WA.

Yahgmour, K. (2003). *Building embedded Linux systems.* CA: O'Reilly.

Jorwekar, S. (2005). *Client server software architecture.*

Leeman, M., Atienza, D., Deconinck, G., De Florio, V., Mendías, J.M., Ykman-Couvreur, C., Catthoor, F., & Lauwereins, R. (2005). Methodology for refinement and optimisation of dynamic memory management for embedded systems in multimedia applications. *Journal of VLSI Signal Processing, 40*(3), 383-396.

RFC 783. (1981). *The TFTP protocol (revision 2).* Retrieved April 6, 2006, from http://www.ietf.org/rfc/rfc0783.txt?number=0783

RFC 1889. (1996). *RTP: A transport protocol for real-time applications.* Retrieved April 6, 2006, from http://www.ietf.org/rfc/rfc1889.txt?number=1889

Shaw, M., & Garlan, D. (1996). *Software architecture: Perspectives on an emerging discipline.* Englewood Cliffs, NJ: Prentice Hall.

Wolf, W. (2005). *Computer as components: Principles of embedded computing system design.* San Francisco: Morgan Kaufmann.

KEY TERMS

Client-Server Architecture: A basic concept used in computer networking, wherein servers retrieve information requested by clients, and clients desplay that information to the user.

Embedded Systems: An embedded system is a special-purpose computer system, which is completely encapsulated by the device it controls. An embedded system has specific requirements and performs pre-defined tasks, unlike a general purpose personal computer.

Embedded Software: Software designed for embedded systems.

Layered Architecture: The division of a network model into multiple discrete layers, or levels, through which messages pass as they are prepared for transmission.

Mobile Devices: Any portable device used to access a network (Internet, for example).

Multimedia Application: Applications that support the interactive use of text, audio, still images, video, and graphics.

Network Protocols: A set of rules and procedures governing communication between entities connected by the network.

Wireless Network: Networks without connecting cables, that rely on radio waves for transmission of data.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 260-264, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.10
Design of an Enhanced 3G-Based Mobile Healthcare System

José Ruiz Mas
University of Zaragoza, Spain

Julián Fernández Navajas
University of Zaragoza, Spain

Eduardo Antonio Viruete Navarro
University of Zaragoza, Spain

Antonio Valdovinos Bardají
University of Zaragoza, Spain

Carolina Hernández Ramos
University of Zaragoza, Spain

Robert S. H. Istepanian
Kingston University, UK

Álvaro Alesanco Iglesias
University of Zaragoza, Spain

José García Moros
University of Zaragoza, Spain

ABSTRACT

An enhanced mobile healthcare multi-collaborative system operating over Third Generation (3G) mobile networks is presented. This chapter describes the design and use of this system in different medical and critical emergency scenarios provided with universal mobile telecommunications system (UMTS) accesses. In these environments, it is designed to communicate healthcare personnel with medical specialists in a remote hospital. The system architecture is based on advanced signalling protocols that allow multimedia multi-collaborative conferences in IPv4/IPv6 3G scenarios. The system offers real-time transmission of medical data and videoconference, together with other non real-time services. It has been optimized specifically to operate over 3G mobile networks using the most appropriate codecs. Evaluation results show a reliable performance over IPv4 UMTS accesses (64 Kbps in the uplink). In the future, advances in m-Health systems will make easier for mobile patients to interactively get the medical attention and advice they need.

INTRODUCTION

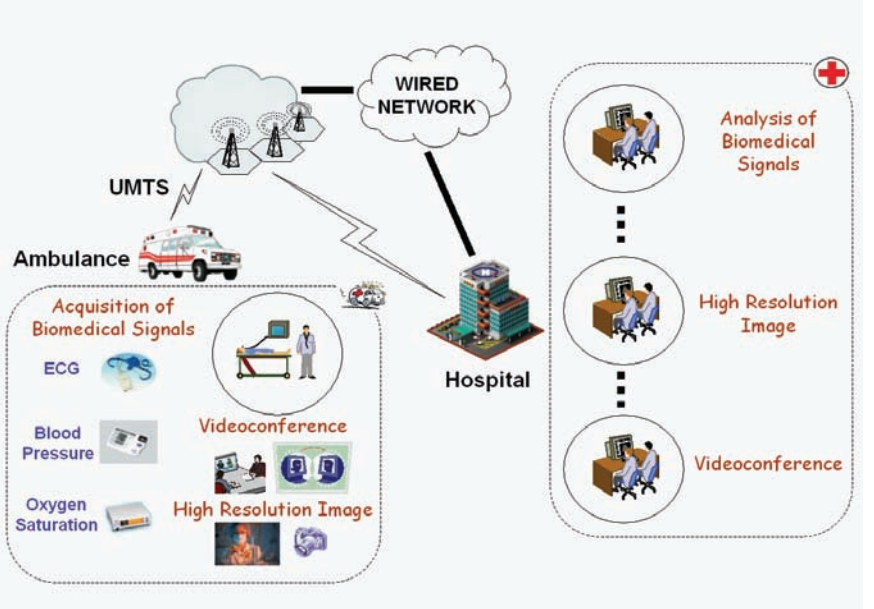
Mobile health (m-health) is an emerging area of telemedicine in which the recent development in mobile networks and telemedicine applications converge. m-health involves the exploitation of mobile telecommunication and multimedia technologies and their integration into new mobile healthcare delivery systems (Istepanian & Lacal, 2003). Wireless and mobile networks have brought about new possibilities in the field of telemedicine thanks to the wide coverage provided by cellular networks and the possibility of serving moving vehicles. One of the first wireless telemedical systems that utilized second-generation (2G) global system for mobile communications (GSM) networks addressed the electrocardiogram (ECG) transmission issues (Istepanian, 2001a). In recent years, several m-health and wireless telemedical systems based on GSM were reported (Istepanian, 2001b), allowing the accomplishment of remote diagnosis in mobile environments, as well as communication to geographic zones inaccessible by wired networks. The recent developments in digital mobile telephonic technologies (and their impact on mobility issues in different telemedical and telecare applications) are clearly reflected in the fast growing commercial domain of mobile telemedical services. A comprehensive review of wireless telemedicine applications and more recent advances on m-health systems is presented in Istepanian, Laxminarayan, and Pattichis (2005).

However, 2G-based systems lack the necessary bandwidth to transmit bandwidth-demanding medical data. The third-generation (3G) universal mobile telecommunications system (UMTS) overcomes limitations of first and second mobile network generations supporting a large variety of services with different quality of service (QoS) requirements. However, this fact makes network design and management much more complex. New applications require networks to be able to handle services with variable traffic conditions keeping the efficiency in the network resources utilization. The UMTS air interface is able to cope with variable and asymmetric bit rates, up to 2 Mbps and 384 kbps in indoor and outdoor environments, respectively, with different QoS requirements such as multimedia services with bandwidth on demand (Laiho, Wacker, & Novosad, 2000). In this kind of scenario, the emergence of 3G mobile wireless networks will permit to extend the use of m-health applications thanks to the provided higher transmission rates and flexibility over previous mobile technologies.

UMTS introduces the IP multimedia core network subsystem (IMS) (3GPP, 2005a), an IPv6 network domain designed to provide appropriate support for real-time multimedia services, independence from the access technologies and flexibility via a separation of access, transport and control. The fundamental reason for using IPv6 is the exhaustion of IPv4 addresses. Support for IPv4 is optional, but since network components require backward compatibility, it is clear that a dual stack configuration (IPv4 and IPv6) must be provided. The IMS uses the session initiation protocol (SIP) as signalling and session control protocol (Rosenberg et al., 2002). SIP allows operators to integrate real-time multimedia services over multiple access technologies such as general packet radio service (GPRS), UMTS or, ultimately, other wireless or even fixed network technologies (interworking multimedia domains). This chapter presents a 3G-based m-health system designed for different critical and emergency medical scenarios, as shown in Figure 1. Several medical specialists in the hospital take part in a multipoint conference with the ambulance personnel, receiving compressed and coded biomedical information from the patient, making it possible for them to assist in the diagnosis prior to its reception.

The 3G system software architecture includes intelligent modules such as information compression and coding, and QoS control to significantly improve transmission efficiency, thus optimizing the use of the scarce and variable wireless channel

Figure 1. Typical medical mobility scenario



bandwidth compared to previous systems (Chu & Ganz, 2004; Curry & Harrop, 1998). Finally, unlike Chu and Ganz (2004), this m-health system follows a multi-collaborative design which supports IPv6/IPv4 interworking, uses SIP as the service control protocol and integrates new real-time multimedia features intended for 3G wireless networks.

3G M-HEALTH SYSTEM ARCHITECTURE

In this section, the 3G m-health system structure is described in detail. The system (see the system components in Figure 2 and the main application graphical user interface (GUI) in Figure 3) has been built using standard off-the-shelf hardware, instead of developing propriety hardware as in Cullen, Gaasch, Gagliano, Goins, and Gunawardane (2001), uses free software and commercially available 3G wireless UMTS cellular data services. In addition, it provides simultaneous transfer of, among other services, videoconference, high-resolution still medical images and medical data, rather than only one media at a time (Kyriacou, 2003; Pavlopoulos, Kyriacou, Berler, Dembeyiotis, & Koutsouris, 1998).

The 3G m-health system consists of different modules that allow the acquisition, treatment, representation and transmission of multimedia information. The modular design of each medical user service allows great flexibility. In addition, there exist other medical user services like chat and electronic whiteboard that allow data exchange in order to guide the operations performed by remote users.

The medical signals module acquires, compresses, codes, represents, and transmits medical signals in real time. The medical signals acquisition devices included are: a portable electrocardiograph that allows the acquisition of 8 real and 4 interpolated leads of the ECG signal, and that follows the standard communication protocol-ECG (SCP-ECG); a tensiometer that provides systolic and diastolic blood pressure values; and a pulsioximeter that offers the blood oxygen saturation level (SpO_2) and the cardiac pulse.

The details of the 3G system architecture are shown in Figure 4 and Figure 5. As it can

Figure 2. 3G m-Health system

Figure 3. m-Health application GUI

be seen, the system comprises of the signalling and session control, medical user services and application control sub-systems, which will be described later.

This architecture allows the 3G system to offer real-time services such as medical data transmission (ECG, blood pressure, heart rate and oxygen saturation), full-duplex videoconference, high-resolution still images, chat, electronic whiteboard, and remote database Web access.

Communication between the remote ambulance personnel and medical specialists is established by means of multipoint multi-collaborative sessions through several network environments capable of supporting the different types of multimedia traffic (Figure 4). The selected con-

Figure 4. 3G network scenario

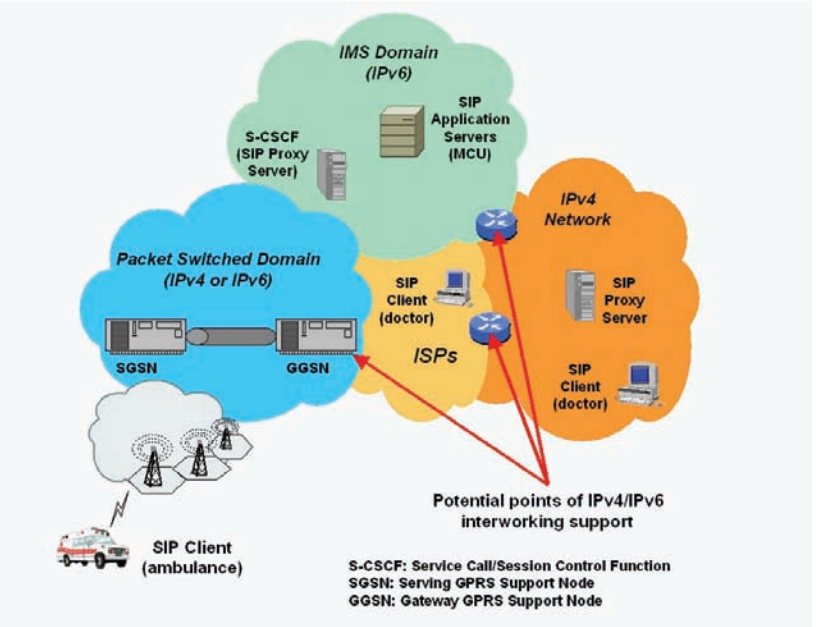
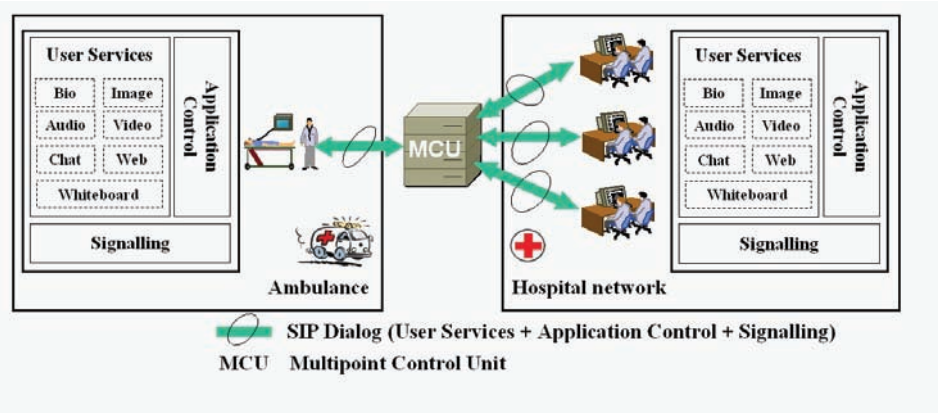


Figure 5. Block diagram of the m-Health system architecture and sub-systems



ference model (tightly coupled conference model (Rosenberg, 2004)) requires the existence of a MCU (multipoint control unit) to facilitate multipoint operation. The MCU maintains a dialog with each participant in the conference and is responsible for ensuring that the media streams which constitute the conference are available to the appropriate participants. The MCU can belong to mobile's home network (SIP application server) or to a external service platform. Furthermore, the m-health system is developed to support IPv4/IPv6 interworking (Wiljakka & Soinien, 2003) and can be integrated in a 3G network scenario (see Figure 4).

The MCU receives the information generated by each participant in the multipoint conference, processes and forwards it to the appropriate destinations. System users and the MCU exchange

information associated with the different services provided (medical user services) and its presentation (application control). Moreover, they exchange information related to communication and service quality management (signalling). Next, a detailed description of the sub-systems is presented.

Signalling and Session Control

The developed signalling allows the exchange of the characteristics associated to the different information flows between system elements and is based on standard protocols that favour interoperability. Signalling tasks, performed by the SIP protocol, begin with the establishment of a SIP dialog with the MCU in which, by means of session description protocol (SDP) messages, the different services are described. In order to do that, each element in the system has a SIP user agent (UA), slightly modified in the MCU to allow the use of multiple simultaneous dialogs.

In addition to session control functions (establishment, management and termination of the multipoint conference), the SIP protocol is also useful for user mobility purposes inside the IMS environment.

Multipoint conference establishment, management and termination is performed by exchanging SIP messages between the different users. When a user connects, he creates a SIP dialog with the MCU, joining the conference. During the conference, SIP messages are exchanged between users and the MCU, varying conference characteristics and therefore allowing its management. In a similar process to that of conference joining, when a user wants to leave it, this fact must be communicated to the MCU with the necessary SIP messages. SIP messages also serve as the mean of transport of SDP messages with the description of medical user services.

The QoS in this system is mainly determined by the characteristics of the UMTS link. Mobile links are very variable, therefore a QoS-monitoring process is required in order to obtain a good system performance. This process is especially important in the MCU because it is there where the QoS-related decisions are taken. When the MCU detects that a particular conference participant needs to modify the characteristics of its multimedia session in order to improve QoS, it renegotiates the corresponding session by sending SIP/SDP messages. Hence, conference participants can modify certain upper-level protocol parameters (codecs used, transmission rates, compression ratios, etc.) in order to adapt the transmitted information to network performance.

The QoS monitoring process is possible thanks to a transport library that provides a uniform interface to send the information generated by medical user services and different QoS estimation tools developed for several types of links. This transport library offers different queuing policies and tools designed to measure the following QoS-related parameters: delay, bandwidth and packet loss rate. Due to the variable nature of wireless links, reception buffers have been properly dimensioned to minimize jitter, delay and packet loss.

Wireless Medical User Services

The medical user services included in the m-health system are associated with information shared in a multi-collaborative environment. Specifically, the system has services to share audio, video, medical data information, high-resolution still images, and graphical and textual information, as well as a Web service that allows remote access to clinical information databases. In addition to these services, there is a service designed to exchange control information (application control), which is discussed later.

Each kind of information is associated with a medical user service and uses a transport protocol and a codec according to its characteristics (see Table 1). Hence, real-time services (audio, video, and medical data information) use the real-time

Table 1. Codec operation modes for 3G real-time wireless medical user services

	CODEC	CODEC RATE
Audio	AMR*	4.75 5.15 5.9 6.7 7.4 7.95 10.2 12.2 (Kbps)
Video	H.263	5 10 (Frames per second)
Biomedical Signals	WT**	5 10 20 (Kbps)

* **Adaptive Multi-Rate**

** **Wavelet Transform**

** Adaptive multi-rate, ** Wavelet transform*

transport protocol (RTP), whereas the rest of the services use the transmission control protocol (TCP). Furthermore, the exchanged information can be very sensitive and requires a secure communication. The 3G m-health system uses an IP security protocol (IPSec) implementation supporting public key certificates in tunnel mode. This protocol ensures private communication of selected services.

The ECG signal is stored both in transmission and reception following the SCP-ECG standard. It is well known that for an efficient transmission an ECG compression technique has to be used. In our implementation, a real-time ECG compression technique based on the wavelet transform is used (Alesanco, Olmos, Istepanian, & García, 2003). This is a lossy compression technique, therefore the higher the compression ratio (lower the transmission rate), the higher the distortion at reception. It is clear that there is a trade-off between transmission rate and received ECG signal quality. From the transmission efficiency point of view, a very low transmission rate is desired but from the clinical point of view, a very distorted ECG is useless. Therefore, there exists a minimum transmission rate to be used so as the transmitted ECG is useful for clinical purposes, which was selected in collaboration with cardiologists after different evaluation tests. The minimum transmission rate used in our implementation (625 bits per second and per ECG lead) leads to a clinically acceptable received ECG signal.

Regarding blood pressure, oxygen saturation, and heart rate, these signals have low bandwidth requirements and, therefore, are not compressed.

The videoconference module captures, sends, and plays audio and video information obtained by means of a Web camera and a microphone. In order to reduce the bandwidth, these data are compressed and coded. The video signal is compressed following the H.263 standard, whereas the audio signal uses the adaptive multi-rate (AMR) codec, recommended for UMTS by the 3G Partnership Program (3GPP) (3GPP 2005b). This module provides the basic functionality for starting, pausing and stopping video signals acquisition and representation, as well as volume control for the microphone (capture) and the speakers (reproduction). Due to the fact that each participant in the conference receives a unique video signal, the system allows the user to select the particular video signal among all the users connected.

The high-resolution still image module obtains high quality images with a charge copled device (CCD) colour camera connected to the computer through an image acquisition card. This module includes options to preview the captured images and modify their main characteristics in real time: brightness, contrast, hue, etc. Captured

images can be stored and transmitted in different formats, with various qualities and compression levels. These images are sent automatically to the electronic whiteboard module of the remote users, allowing to select and mark fixed areas in a multi-collaborative fashion to facilitate a diagnostic clinical procedure.

Application Control

The MCU forwards the information generated by each medical service according to the presentation spaces defined using the control service. Each medical service has a presentation space associated with it that defines the way in which the information has to be transferred and its destination. The MCU simply forwards the information it receives, but has a special treatment for the audio, video, medical data, and control services. Regarding the Audio service, the MCU decodes the signal of each user, mixes it with the decoded signal of the other conference participants and codes the result in order to transfer a unique audio signal to each user (Figure 6). On the other hand, the MCU only forwards one video signal to each conference participant. The particular video signal forwarded to each user is selected by using the control service. Finally, the medical data service is similar to the video service: medical data are only generated in the remote location, whereas the other conference participants can only receive them.

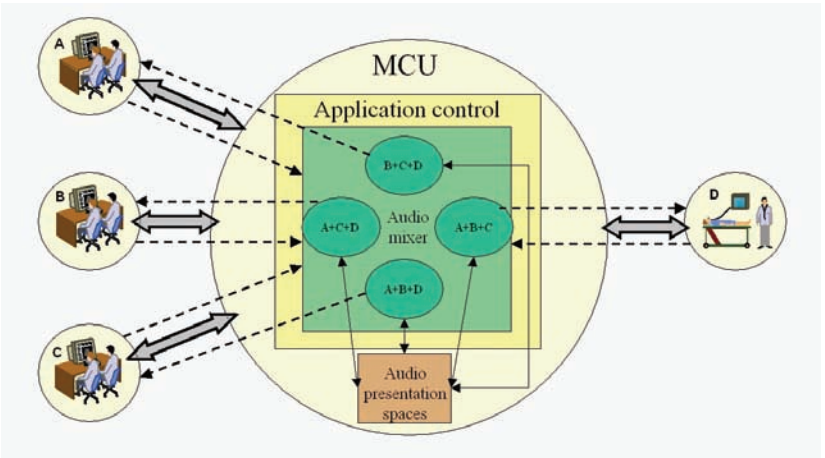
The 3G m-health system was adapted to the critical and emergency medical scenarios. In the first stages of its design, user requirements and functional specifications were established in collaboration with medical specialists, in order to create a portable and modular m-health system that could be easily integrated in any environment, using any underlying network technology capable of supporting IP multimedia services

3G M-HEALTH SYSTEM PERFORMANCE

In order to measure the 3G m-health system performance, several tests have been carried out using the system over 64 Kbps (at IP level) UMTS accesses. Table 2 presents the results about average IP-level bandwidth used by real-time network services.

As it can be observed, considering more audio samples per network packet reduces the used bandwidth, since transmission efficiency (information carried by each packet to total packet

Figure 6. Application control (audio presentation spaces)



size ratio) is increased. However, there is a limit in the number of audio samples per packet that can be used because more audio samples per packet yield more audio delay. For example, a more efficient transmission mode including four audio samples per packet every 80 ms causes four times the delay including only one audio sample per packet every 20 ms. Moreover, if an audio packet is lost, all the audio samples carried by it are lost and, therefore, a reduced number of audio samples per packet is more suitable to error-prone environments. Regarding the video user service, it is worth noting that the bandwidth shown in Table 2 can vary substantially with the movement of the video scene captured. Finally, the medical data service adapts well to the codec rate specified because medical data frame sizes are long enough to obtain a good transmission efficiency.

As it can be checked, the total bandwidth consumed by all real-time medical user services fits in a 64 Kbps UMTS channel, even when the most bandwidth-consuming codec rates and the lowest transmission efficiencies are used. If all medical user services are used (including non real-time services), lower codec rates should be selected. Thus, according to the previous discussions, the codec operation modes selected in this m-health system have been those highlighted in Table 2, achieving a reasonable trade-off between bandwidth, transmission efficiency, delay and loss ratio.

NEXT-GENERATION OF M-HEALTH SYSTEMS

It is evident that organizations and the delivery of health care are being underpinned by the advances in m-health technologies. In the future, home medical care and remote diagnosis will become common, check-up by specialists and prescription of drugs will be enabled at home and virtual hospitals with no resident doctors will be realized.

Hence, the deployment of emerging mobile and wireless technologies will face new challenges: inter-operability between heterogeneous networks (fourth-generation, 4G) and smart medical sensor design integrating sensing, processing, communications, computing, and networking.

With the aid of wireless intelligent medical sensor technologies, m-health can offer healthcare services far beyond what the traditional telemedical systems can possibly provide. The individual sensors can be connected wirelessly to a personal monitoring system using a wireless body area network (WBAN) and can be integrated into the user's clothing, providing wearable and ubiquitous m-health systems. A typical scenario

Table 2. Average IP-level bandwidth used by real-time user services

	Operation Mode		IP Bandwidth (Kbps)
	Samples/packet	Codec rate (Kbps)	
Audio	1	4.75	21.2
	1	12.2	28.8
	3	**4.75**	**10.5**
	3	12.2	18.1
	Frames per second		
Video	**5**		**16**
	10		24
	Bit Rate		
Biomedical Signals	**5**		**5.3**
	10		10.3

comprises of a WBAN system that communicates with cardiac implantable devices (pacemakers, defibrillators, etc.) and that is linked to the existing 3G infrastructure, achieving "Health Anytime, Anywhere" (Figure 7).

It is expected that 4G will integrate existing wireless technologies including UMTS, wireless LAN, Bluetooth, ZigBee, Ultrawideband, and other newly developed wireless technologies into a seamless system. Some expected key features of 4G networks are: high usability, support for multimedia services at low transmission cost and facilities for integrating services. 4G advances will make easier for mobile patients to interactively get the medical attention and advice they need. When and where is required and how they want it regardless of any geographical barriers or mobility constraints.

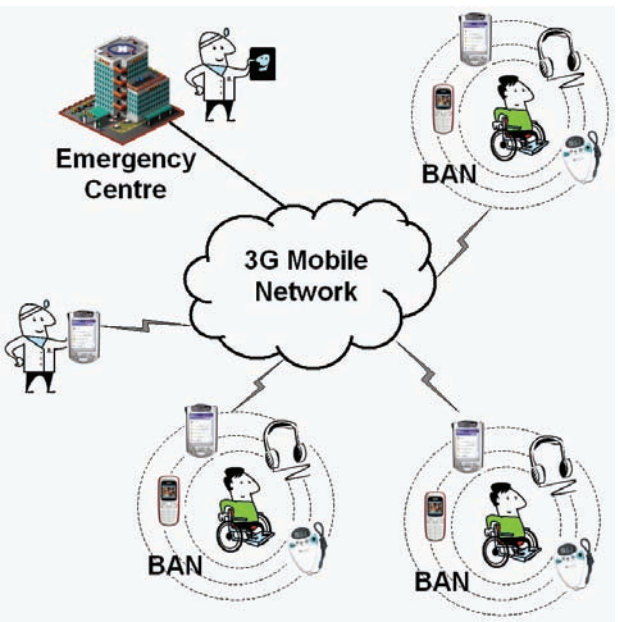
The concept of including high-speed data and other services integrated with voice services is emerging as one of the main points of future telecommunication and multimedia priorities with the relevant benefits to citizen-centered healthcare systems. The new wireless technologies will allow both physicians and patients to roam freely, while maintaining access to critical patient data and medical knowledge.

CONCLUSION

This chapter has presented a feasible 3G-based m-health system targeted specifically for critical and emergency medical scenarios. The system architecture is based on 3G networks and advanced signalling protocols (SIP/SDP) that allow the integration of real-time multimedia services over multiple access channels that support IPv4 and IPv6 interworking depending on current commercial UMTS releases.

The system has the following features: simultaneous transmission of real-time clinical data (including ECG signals, blood pressure, and blood oxygen saturation), videoconference, high-resolution still image transmission, and other facilities such as multi-collaborative whiteboard, chat, and Web access to remote databases. The system has been optimized specifically to operate

Figure 7. Typical WBAN-3G scenario



over 3G mobile networks using the most appropriate codecs. Evaluation results show a reliable performance over UMTS accesses (64 Kbps in the uplink).

Home telecare and chronic patient telemonitoring are other application areas in which this m-health system can be used, thus further work is currently undergone to adapt it and to evaluate its performance in each particular scenario.

REFERENCES

3GPP TS 23.228 V6.8.0. (2005). IP Multimedia Subsystem (IMS); Stage 2 (Release 6).

3GPP TS 26.235 V6.3.0. (2005). Packet switched conversational multimedia applications; Default codecs (Release 6).

Alesanco, A., Olmos, S., Istepanian, R. S. H., & García, J. (2003). A novel real-time multilead ECG compression and de-noising method based on the wavelet transform. *Proceedings of IEEE Computers Cardiology* (pp. 593-596). Los Alamitos, CA: IEEE Comput. Soc. Press.

Chu, Y., & Ganz, A. (2004). A mobile teletrauma system using 3G networks. *IEEE Transactional Information Technology in Biomedicine, 8*(4), 456-462.

Cullen, J., Gaasch, W., Gagliano, D., Goins, J., & Gunawardane, R. (2001). *Wireless mobile telemedicine: En-route transmission with dynamic quality-of-service management.* National Library of Medicine Symposium on Telemedicine and Telecommunications: Options for the New Century.

Curry, G. R., & Harrop, N. (1998). The Lancashire telemedicine ambulance. *Journal Of Telemedicine Telecare, 4*(4), 231-238.

Istepanian, R. S. H., Kyriacou, E., Pavlopoulos, S., & Koutsouris, D. (2001). Wavelet compression methodologies for efficient medical data transmission in wireless telemedicine system. *Journal of Telemedicine and Telecare, 7*(1), 14-16.

Istepanian, R. S. H., Laxminarayan, S., & Pattichis, C. S. (2005). *M-Health: Emerging mobile health systems.* New York: Springer. To be published.

Istepanian, R. S. H., & Lacal, J. C. (2003). Emerging mobile communication technologies for health: Some imperative notes on m-health. *Proceedings of the 25th Silver Anniversary International Conference of the IEEE Engineering in Medicine and Biology Society 2* (pp. 1414-1416).

Istepanian, R. S. H., Woodward, B., & Richards, C. I. (2001). Advances in telemedicine using mobile communications. *Proceedings of the IEEE Engineering Medicine and Biology Society 4* (pp. 3556-3558).

Kyriacou, E., Pavlopoulos, S., Berler, A., Neophytou, M., Bourka, A., Georgoulas, A., Anagnostaki, A., Karayiannis, D., Schizas, C., Pattichis, C., Andreou, A., & Koutsuris, D. (2003). Multipurpose healthcare telemedicine systems with mobile communication link support. *BioMedical Engineering OnLine, 2*(7).

Laiho, J., Wacker, A., & Novosad, T. (2000). *Radio network planning and optimization for UMTS*. New York: Wiley.

Pavlopoulos, S., Kyriacou, E., Berler, A., Dembeyiotis, S., & Koutsouris, D. (1998). A novel emergency telemedicine system based on wireless communication technology–AMBULANCE. *IEEE Trans. Inform. Technol. Biomed, 2*, 261-267.

Rosenberg, J. (2004). *A framework for conferencing with the session initiation protocol.* Internet draft. Work in progress.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). *SIP: Session initiation protocol* (IETF RFC 3261).

Wiljakka, J., & Soinien, J. (2003). Managing IPv4-to-IPv6 transition process in cellular networks and introducing new peer-to-peer services. *Proceedings of IEEE Workshop on IP Operations and Management* (pp. 31-37).

KEY TERMS

4G: The fourth-generation (4G) is the continuation of the first, second, and third generations of mobile networks. It is a wireless access technology that provides high-speed mobile wireless access with a very high data transmission speed (2-20 Mbps) and enables users to be simultaneously connected to several wireless access technologies and seamlessly move between them in an all-IP environment. These access technologies can be any existing or future access technology. Smart antennas, low power consumption, and software-defined radio terminals will also be used to achieve even more flexibility for the user of 4G systems.

IMS: The IP multimedia subsystem (IMS) is a new open and standardized framework, basically specified for mobile networks, for providing Internet protocol (IP) telecommunication services. It offers a next generation network (NGN) multimedia architecture for mobile and fixed services, based on the session initiation protocol (SIP), and runs over the standard IP. It is used by telecom operators in NGN networks (combining voice and data in a single packet switched network), to offer network controlled multimedia services. The aim of IMS is not only to provide new services but to provide all the services, current and future, that the Internet provides. In addition, users have to be able to execute all their services when roaming as well as from their home networks. To achieve these goals the IMS uses IETF protocols. This is why the IMS merges the Internet with the cellular world; it uses cellular technologies to provide ubiquitous access and Internet technologies to provide new services.

IPv6: Internet protocol version 6 (IPv6) is the next generation protocol designed by the Internet Engineering Task Force (IETF) to replace the current version of the Internet protocol, IP version 4 (IPv4). Today's Internet has been using IPv4 for twenty years, but this protocol is beginning to become outdated. The most important problem of IPv4 is that there is a growing shortage of IPv4 addresses, which are needed by all new machines added to the Internet. IPv6 is the solution to several problems in IPv4, such as the limited number of available IPv4 addresses, and also adds many improvements to IPv4 in other areas. IPv6 is expected to gradually replace IPv4, with the two coexisting for a number of years during a transition period.

QoS: ITU-T recommendation E.800 defines the term quality of service (QoS) as "the collective effect of service performance which determines the degree of satisfaction of a user of the service." Service performance comprises of very different parts (security, operability, etc), so the meaning of this term is very broad. In telecommunications, the term QoS is commonly used in assessing whether a service satisfies the user's expectations. QoS evaluation, however, depends on functional components and is related to network performance via measurable technical parameters. A QoS-enabled network has the ability to provide better service (priority, dedicated bandwidth, controlled jitter, latency, and improved loss characteristics) to selected network traffic over various technologies.

SIP: The session initiation protocol (SIP) is a signalling protocol developed by the IETF intended for setting up multimedia communication sessions between one or multiple clients. It is currently the leading signalling protocol for voice over IP, and is one of the key components of the IMS multimedia architecture. The most important functions of SIP include name mapping and redirection (user location), capabilities negotiation during session setup, management of

session participants and capabilities management during the session.

Telemedicine: Telemedicine can be defined as the rapid access to shared and remote medical expertise by means of telecommunications and information technologies, no matter where the patient or the relevant information is located. Any application of information and communications technologies which removes or mitigates the effect of distance in healthcare is telemedicine. The terms e-health and tele-health are terms often interchanged with telemedicine.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 521-533, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.11
Service Provisioning in the IP Multimedia Subsystem

Adetola Oredope
University of Essex, UK

Antonio Liotta
University of Essex, UK

INTRODUCTION

The IP multimedia subsystem (IMS) specifies a service-centric framework for converged, all-IP networks. This promises to provide the long awaited environment for deploying technology-neutral services over fixed, wireless, and cellular networks, known as third generation (3G) networks. Since its initial proposal in 1999, the IMS has gone through different stages of development, from its initial Release 5 up to the current Release 7.

The IMS is also known as a domain for easily integrating and customising different services offering a new range of applications which aim at increasing the end-user experience. The IMS architecture will deploy both current and future Internet services on multi-access, 3G Networks (Camarillo & Garcâia-Martâin, 2006). These services include voice over IP (VoIP), instant messaging, video conferencing, online gaming, push-to-talk, and whiteboard sharing, to mention a few.

Service provisioning in the IMS offers the required mechanisms for execution, control, and integration of these services, including better and more efficient billing, accounting, quality of service (QoS), and interoperability across different administrative domains. All present and future Internet services will be deployed on a multi-access, all-IP overlay network via the IMS.

This version is based on the session initiation protocol (SIP) (Rosenberg et al., 2002) for its core functionality. SIP is used to create, modify and terminate sessions within the framework. SIP is also used to deliver sessions descriptions between SIP entities using the session description protocol (SDP) (Handley & Jacobson, 1998). The IMS provides an open interface for third-party services, such as the open service access (OSA) (3GPP-TS29.198, 2001) to be easily integrated, providing an opportunity for end-users to expe-

rience rich and personalized services (Liotta et al., 2002).

In this paper, the second section provides a brief description of the IMS network and service architectures required for deploying advanced IP services over converged networks. It also explains the IMS services capabilities standardization efforts and collaboration between the Third Generation Partnership Project (3GPP), 3GPP2, and the Open Mobile Alliance (OMA). The third section describes service provisioning in the IMS, providing the key elements for service execution (session and presence management), service quality control (quality control and billing), and service integration (open service access and converged billing), which are basic enablers for IMS service provisioning. The fourth section describes the future trends for service provisioning, outlining both academic and industrial opportunities in these areas. Finally, in the fifth section we draw key conclusions about service provisioning in the IMS.

IMS SERVICE ARCHITECTURE

IMS is based on a service oriented architecture (SOA) in which different services and applications from different vendors can easily communicate and be integrated to develop new services. In the IMS, service capabilities are standardised and not services. To help understand the IMS Service architecture, a brief description of the IMS network architecture (3GPP-TS23.002, 2006) and the standardisation service capabilities are first provided.

IMS Network Architecture

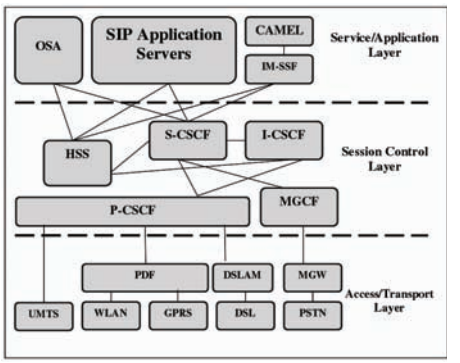
The IMS network architecture is built on a layered approach with open and standard interfaces to allow for seamless interconnection of standardized functions. In the IMS, nodes are not standardized, allowing vendors to combine as many functions as they wish into a particular node (Camarillo & Garcâia-Martâin, 2006). The horizontal layered approach of the IMS allows for lower layers to be transparent to the upper layers, enabling operators and service developers to use different underlying networks. This advantage allows interoperability and roaming. The IMS network architecture is divided in three basic layers as explained below (Figure 1).

A full description of the IMS network architecture can be found in 3GPP-TS23.002, 2006). From Figure 1, the Access Layer is an access-independent interface that allows users to connect to the IMS network via existing fixed or wireless networks, The session control layer is made up of SIP servers and proxies for controlling and managing sessions within the IMS. This is made up of the Proxy CSCF (P-CSCF), interrogating CSCF (I-CSCF), and the serving CSCF (S-CSCF). The service/application layer is also made up of different application servers for the execution of various IMS services and the provision of end user service logic.

Standardisation of Service Capabilities

Service capabilities are mechanisms needed to realise services within the IMS network and under the network control (3GPP-TS22.105, 2006). They are usually standardised to allow for service differentiation and system continuity. Service ca-

Figure 1. The IMS network architecture



pabilities include descriptions for bearer services, teleServices, and supplementary services.

The standardisation of service capabilities exists in other systems such as GPRS, but in the IMS, 3GPP, 3GPP2, and OMA are responsible for the standardisation of the service capabilities.

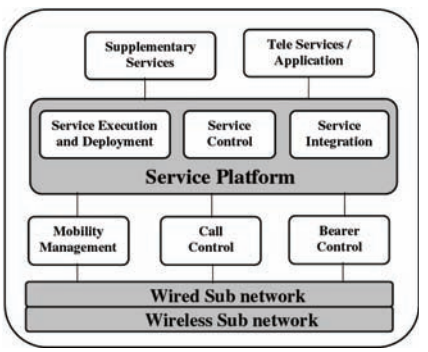
Some of OMA works are based on the IMS platform, and this leads to the proposals to extend the IMS in order to support additional requirements or service. This is the basis of the agreement between OMA, 3GPP, and 3GPP2 in which OMA generates the requirements for the IMS and the 3GPP and 3GPP2 extend the IMS to support the specified requirement (Camarillo & Garcâia-Martâin, 2006).

Due to this collaboration, OMA and 3GPP work on similar specifications at times and they aim to have compatible specifications. Examples of similar specifications being worked on by OMA and 3GPP include presence, messaging, and push-to-talk/push-over-cellular (PTT/PoC).

IMS Service Architecture

Based on the standardisation of service capabilities in Section 2.2, an IMS service architecture is developed to allow easy deployment and management of services. From the diagram, the Wired and Wireless sub networks provide interfaces that allow the IMS services to be transparent to the access network. These interfaces are explained below.

Figure 2. IMS service architecture (3GPP-TS22.105, 2006)



Bearer Control

Bearers allow information to be efficiently transferred to teleservices and applications via sub networks providing different levels of quality of service (QoS). They are usually classified with parameters such as “throughput,” “delay tolerance,” “maximum bit error rate,” “symmetry,” etc. (3GPP-TS22.105 2006). The bearer control plays a major role during interworking of functions and content adaptation.

Call Control

Call control allows the network operator to have total control of all information in the IMS network. This is achieved via a SIP proxy server known as a back-to-back user agent (B2BuA) placed in the signal and media path of end-to-end services between users. The B2BUA is based on an architecture that modifies the SIP messages (headers and bodies), generates and responds to requests (Camarillo & Garcâia-Martâin, 2006). The B2BUA allows the network operator full control of the traffic in the network, which in return allows for proper session management, resource reservation, internetwork billing, signalling interworking, protocol adaptation, and provision of integrated services. In the IMS, the B2BUA plays the major role of interconnecting the IMS with external networks (3GPP-TR29.962, 2005).

Mobility Management

Mobility is a key issue in the IMS because a large percentage of terminals in the network are mobile devices such as 3G phones, PDAs, laptops, and portable games. In fact, one of the most important challenges IP networks face is mobility support. Mobile IPv4 (MIPv4) (Perkins, 2002) and Mobile IPv6 (MIPv6) (Johnson, Perkins, et al., 2004) are the result of IETF mobility support efforts. These protocols enhance the network layer, so that IP hosts can change location while retaining their

communicating sessions. Due to the limitation of Mobile IP, the IMS manages mobility through GPRS, providing layer two tunnelling mechanisms. This allows for a successful handover to be made without the session or application breaking. Terminal Mobility also allows for roaming, which is led by the CSCF and the HSS to allow the access to the subscribed services either in a home or visited network (Camarillo & Garcâia-Martâin, 2006).

SERVICE PROVISIONING IN THE IMS

Why is IMS Service Provisioning Important?

The IMS aims to provide specialized end-user experience, better billing systems, enhanced security mechanisms, better mobility management (roaming), improved quality of service, and, most importantly, the integration of various services which cannot be found in present 3G networks. The IMS is viewed as a service-centric platform for endless development of new, rich, and exciting services. The IMS has put in place various mechanisms for the execution, control, and integration of theses services as shown in Figure 2. This process is known as service provisioning as explained in the following sections.

Service Execution and Deployment

The services offered in the IMS are hosted on application servers whereby some application servers may host multiple services. When a user registers with the network, the subscriber service profile (SSP) of the user is downloaded by the CSCF from the HSS. The SSP contains unique user-specific information such as:

- The services needed to be executed and the information on the order in which multiple services should be executed
- The location information of the application servers required to execute the services, and the order in which multiple services should be executed when they are on the same application server

The SSP information is used to deliver user-specific or personalized services to the end-users. These services are built on the various service enablers as explained below.

Presence Services

The presence service (PS) is the foundation of service provisioning in the IMS because it allows the collection of various forms of information on the characteristics of the network, users, and terminals. This service enabler allows users to subscribe for certain information (watchers) about other elements (presentities). Presentities can then decide to publish this information, which could include capabilities of terminals, locations, communication addresses, or user availability. Moreover, the IMS can also make the published information available to other services to allow for easy creation of new services and integration with existing services. For example, the IMS terminal plays the role of both a watcher and a presence user agent (PUA). The application server in the home network also plays the role of the presence server (PS), also known as presence agent (PA).

In order to allow for proper service execution in the IMS, the PA needs to get all relevant information from the necessary elements in the network, such as the home location register (HLR) in circuit networks, the gateway GPRS support node (GGSN) in GPRS networks, and the S-CSCF in the IMS network. This allows for proper service logics to be computed in order to deliver personalized services to the user.

Session Management

Sessions management is based on creation, modification, integration, and modification of conventional SIP messages as described in RFC-3261 (Rosenberg, Schulzrinne, et al., 2002). This allows end-users and operators to manage different sessions in the IMS, and also create new services. The CSCF (Figure 1) is at the heart of session management routing. The CSCF is basically a SIP server, and is divided in three major categories: P-CSCF, I-CSCF, and S-CSCF, as described earlier in the second section. IMS also supports SIP extensions in which "required" and "supported" headers are used. This also allows for the creation of new services such as presence (Roach, 2002) and instant messaging (Campbell, Rosenberg, et al., 2002).

The IMS terminal can send a REGISTER request via the P-CSCF to the I-CSCF and then the S-CSCF. If the registration is successful, a 200 OK message is replied back to the terminal via the same route. Once registration is complete, sessions to User Agents can be easily managed by invoking new or existing services using various SIP messages such as INVITE, SUBSCRIBE, NOTIFY, and PRACK to mention a few.

Application Servers and Gateways

As described earlier in the second section, application servers are used to host and execute the services used in the IMS network. An application server can host more than a single service. There are three different application servers: namely, the SIP application servers for IMS services, the open service access-service capability server (OSA-SCS), a gateway to execute OSA in the IMS, and the IMS switching function (IM-SSF)—a gateway between the IMS and CAMEL, which is deployed in GSM networks.

Gateways, on the other hand, are used to interconnect other standardized formats to the standards supported in the IMS. Examples include: the Media Gateway, which interfaces the media plane in the circuit-switched network, converting between PCM and RTP. Also, the media gateway controller function (MGCF) is used as call-control protocol conversion between SIP and ISUP (integrated services digital network user part). The latter defines the procedures used to setup, manage, and release trunk circuits that carry voice and data calls over the public switched telephone network (PSTN).

Group List Management

This enabler is part of the presence service (PS). It allows users to create and manage network-based groups using definitions of the services deployed in the IMS. It allows the users to have access to information on, and receive notifications about, the groups. This is used in the development of buddy lists, contact lists, blogs, public/private chat groups, and new services where identities are required. This enabler can also be used for roaming purposes, where the users need access to the usual customized services.

IMS Service Control

Service control (Figure 2) plays a major role in IMS service provisioning due to the fact that it allows the synchronization needed for session establishment and quality of service (QoS) management. In this way, the end-users can have a predictable experience at a reasonable charge as compared to present 3G networks. Service control also allows the service platform to be reusable by allowing proper control and management of complex functions such as service filtering, triggering, and interaction. The main enablers for service control are: quality control, mobility management, security, and billing, as described below.

Quality Control

Quality control in the IMS is described in terms of end-to-end QoS, aiming at a predictable user experience. All steps are taken to ensure that all necessary policies are enforced to guarantee the assigned QoS (Dong, 2005).

In order to allow end-to end QoS, QoS is managed within each domain. The IMS supports various end-to-end QoS models in which the basic reservation protocols are used. The terminals can use either the Integrated Services—including the resource reservation protocol (RSVP)—or differentiated services (DiffServ) codes, as long as they are mapped to the DiffServ codes in the network (Borosa, Marsic, et al., 2003). The link-layer resource reservation is made over a policy decision point (PDP) context and is assigned to the appropriate DiffServ code point (DSCP), which is possibly in the same domain, and is then sent into the DiffServ-enabled Network.

Security

Security in the IMS is based on the IP security protocol (IPSec) (Kent & Atkinson, 1998) in which extensions have been created to manage IPSec security associations for SIP in the IMS. IPSec is also the preferred choice for IPv6. It is an extension of IP which allows security to be removed from the network and placed on the endpoints by applying encryption to either the entire IP payload (tunnel mode) or only to the upper-layer protocols of the IP payload (transport mode). IPSec allows for security services such as data integrity protection, data origin authentication, anti-replay protection, and confidentiality, also offering protection to the upper layers. This allows the security mechanism in the IMS to be divided into Access Security (for users and the network) and Network Security (for the protection of traffic between network elements).

Billing

The IMS is built on multimedia messages requiring large file transfers to ensure good quality. By contrast to the billing approach used in conventional Telecom systems, in the IMS the end-users are charged for the services offered rather than by the bytes transferred. This allows, for example, chat sessions to be charged based on duration rather than by the amount of messages transferred. Each session carries a unique IMS charging identifier (ICID) per session and also an inter-operator identifier (IOI) defining the originating and termination networks. There are basically two categories of billing, which are offline for post-paid users, and online charging for prepaid users.

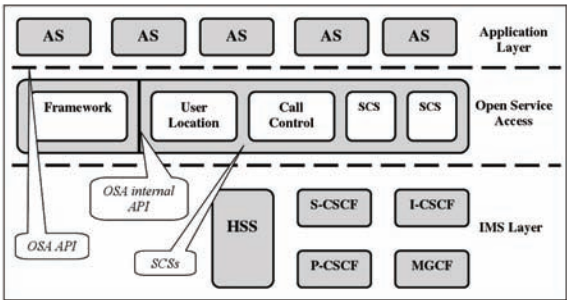
IMS Service Integration

Service integration in the IMS is the element of service provisioning that allows new and rich services to be easily developed. This also allows for an inexpensive integration of both real time and non-real time services to easily construct more complex and attractive "bundled services." The basic enablers for service integration are explained below.

Open Service Architecture (OSA)

The OSA is the defined architecture that allows operators and third-party service providers to make use of the underlying network functionalities through the standardised open standard interface as described in the second section. The OSA also makes the application independent of the underlying network technology (Figure 3) (3GPP-TS23.127 2002). Service capability servers (SCS) provide the application with service capability features (SCF), which are abstractions from the underlying network functionality. SCS implement SCF and interact with the core network. Example SCF are call control, session control, terminal

Figure 3. OSA architecture



capability, charging, account management, and user location.

The framework provides applications with basic mechanisms that enable them to make use of service capabilities in the network. Example framework functions include service discovery and registration, authentication, trust and security management, and service factory.

The main aim of OSA is to provide an extensible and scalable architecture that allows for inclusion of new service capability features and SCS with a minimum impact on the applications, using the OSA interface.

While SCS provide a first, low-level of support aimed at facilitating application development, the application layer offers a second, higher-level of support aimed at speeding up application development. Applications are implemented in one or more application servers (AS), which hide distributional aspects and offer a standard programming environment, application-oriented APIs (e.g., JAIN), and reusable components (e.g., Java Beans) that realise SCF functionality.

Converged Billing

The concept of converged billing allows the operator to charge on the basis of service, content, volume, or an integration of the services in a prepaid or post-paid manner (the modes of payment were discussed under billing earlier on). Converged billing also allows the user to use the operator as their converging point for all their billing, even if some of their services are offered by third-party service providers. This is possible via the single sign on feature in the IMS that allows the user to register with the network, and does not need any form of re-authentication for subsequent services. This gives users the assurance that all their billing is handed by a centrally secure source and gives them enough confidence to spend on the services.

FUTURE TRENDS

The IMS core network architecture is being deployed by various vendors and operators. A lot of end-to-end interoperability testing has been carried out, for instance by Nokia and NEC (Nokia & NEC, 2004). IBM and Swisscom have conducted proof-of-concept tests (IBM, 2006). Also, new services are staring to appear, for instance the proposed mobile gaming architecture (Akkawi, Schaller, et al., 2004) and the platform for mobile TV (Faria, Henriksson, et al., 2006). There also are various tools available to develop IMS services. An example is the Java API for integrated networks (JAIN) in which various expert groups provide reusable APIs that are easily assessed.

The enormous interest surrounding the IMS indicates that this may really become the standard framework for deploying advanced, ubiquitous services over converged, all-IP networks, i.e., the "domain of services." However, not all research areas have been fully explored yet, especially in the area of distributed session management. Many approaches have been developed to eliminate (or reduce) the number of centralized servers from the IMS, aiming at a better level of scalability. Another promising area of research is looking at the deployment of peer-to-peer services in the IMS (Liotta, 2005). This requires coming up with solutions as to how SIP sessions can be managed via distributed signalling protocols (Bryan, 2005).

CONCLUSION

Service provisioning in the IMS enables network operators, service providers, and end-users to benefit from the advantages of an open architecture which relies on a solid framework for service execution, control, and integration. These provisioning enablers include presence and location services, quality control, security, and billing (to mention a few).

The IMS as a whole has great potential and has been developing rapidly over the years, offering a solution to the all-IP vision of rich, multi-access multimedia services accessible anywhere, at any time, with the required quality, and at the right price.

The IMS is now widely considered as the future service-centric platform of preference, although several researchers are still arguing that the benefits of IMS come with a high cost linked to its layered approach (complexity). Only a large-scale deployment of the IMS and a wide adoption by operators and service providers will allow a full assessment of its benefits and shortcomings. These will hopefully be unveiled in the next few years.

REFERENCES

3GPP-TR29.962 (2005). *Signalling interworking between the 3GPP profile of the Session Initiation Protocol (SIP) and non-3GPP SIP usage (Release 6).* (3GPP TR 29.962).

3GPP-TS22.105 (2006). *Services and service capabilities.* (3GPP-TS22.105 version 8).

3GPP-TS23.002 (2006). *Network architecture (Release 7).* (3GPP-TS23.002).

3GPP-TS23.127 (2002). *Virtual home environment/Open service access.* (3GPP-TS23.127).
3GPP-TS29.198 (2001). *Open service architecture (OSA).* (3GPP TS 29.198).

Akkawi, A., et al. (2004). *Networked mobile gaming for 3G-networks.* Entertainment Computing, ICEC 2004, 457-467.

Borosa, T., B. Marsic, et al. (2003). *QoS support in IP multimedia subsystem using DiffServ.* Vol. 2. 669-672.

Camarillo, G., & Garcâia-Martâin, M.A. (2006). *The 3G IP multimedia subsystem (IMS): Merging the Internet and the cellular worlds.* Chichester, West Sussex; Hoboken, NJ: J. Wiley.

Campbell, B., J., et al. (2002). *Session initiation protocol (SIP) extension for instant messaging.* Internet Engineering Task Force (IETF RFC 3428).

Dong, S. (2005). *End-to-end QoS in IMS enabled next generation networks.* WOCC, 28.

Faria, G., et al. (2006). DVB-H: digital broadcast services to handheld devices. *Proceedings of the IEEE, 94*(1), 194-209.

Handley, M., & Jacobson, V. (1998). *SDP: Session description protocol.* Internet Engineering Task Force (IETF RFC 2327).

IBM. (2006). *Service architecture for 3GPP IP multimedia subsystem–The IBM and Swisscom proof-of-concept experience*. Retrieved 31 July, 2007 from http://www-03.ibm.com/industries/telecom/doc/content/bin/swisscomm_02.09.06a.pdf

Johnson, D., et al. (2004). *Mobility support in IPv6*. Internet Engineering Task Force (IETF RFC 3775).

Kent, S., & Atkinson, R. (1998). *Security architecture for the Internet protocol.* Internet Engineering Task Force (IETF RFC 2401).

Liotta, A., et al. (2002). *Delivering service adaptation with 3G Technology,* 108-120.

Nokia and NEC (2004). Nokia and NEC successfully test interoperability of IP multimedia

subsystem (IMS) for richer communications. Retrieved 31 July, 2007 from http://press.nokia.com/PR/200409/960657_5.html

Perkins, C. (2002). *IP mobility support for IPv4.* Internet Engineering Task Force (IETF RGC 3344).

Roach, A. B. (2002). *Session initiation protocol (SIP)-Specific event notification*. RFC 3265—Internet Engineering Task Force (IETF RFC 3265).

Rosenberg, J., et al. (2002). *SIP: Session initiation protocol.* Internet Engineering Task Force (IETF RFC 3261).

KEY TERMS

Application Servers: Application servers are used to host and execute the services used in the IMS network.

IMS Service Provisioning: Service provisioning in the IMS offers the required mechanisms for execution, control, and integration of services in the IMS.

IMS Session Management: Sessions management is based on creation, modification, integration, and modification of conventional SIP messages.

Open Service Architecture (OSA): The OSA is the defined architecture that allows operator and third party applications to make use of the underlying network functionalities through standardized, open interface.

Presence Service: This allows for the collection of various forms of information on the characteristics of the network, users, and terminals. Users can subscribe or be notified of this information.

Services Capabilities: These are standardised mechanisms needed to realise services within the IMS network and under the network control.

Service Integration: This is the integration of both real-time and non-real time services to more complex and attractive "bundled services."

This work was previously published in Encyclopedia of Internet Technologies and Applications, edited by M. Freire and M. Pereira, pp. 525-531, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.12
Adaptive Transmission of Multimedia Data over the Internet

Christos Bouras
Research Academic Computer Technology Institute and University of Patras, Greece

Apostolos Gkamas
Research Academic Computer Technology Institute and University of Patras, Greece

Dimitris Primpas
Research Academic Computer Technology Institute and University of Patras, Greece

Kostas Stamos
Research Academic Computer Technology Institute and University of Patras, Greece

INTRODUCTION

Internet is a heterogeneous network environment and the network resources that are available to real time applications can be modified very quickly. Real time applications must have the capability to adapt their operation to network changes. In order to add adaptation characteristics to real time applications, we can use techniques both at the network and application layers. Adaptive real time applications have the capability to transmit multimedia data over heterogeneous networks and adapt media transmission to network changes. In order to implement an adaptive multimedia transmission application, mechanisms to monitor the network conditions, and mechanisms to adapt the transmission of the data to the network changes must be implemented.

Today, the underlying infrastructure of the Internet does not sufficiently support quality of service (QoS) guarantees. The new technologies, which are used for the implementation of networks, provide capabilities to support QoS in one network domain but it is not easy to implement QoS among various network domains, in order to provide end-to-end QoS to the user. In

addition, some researchers believe that the cost for providing end-to-end QoS is too big, and it is better to invest on careful network design and careful network monitoring, in order to identify and upgrade the congested network links (Diot, 2001).

In this article, we concentrate on the architecture of an adaptive real time application that has the capability to transmit multimedia data over heterogeneous networks and adapt the transmission of the multimedia data to the network changes. Moreover in this article, we concentrate on the unicast transmission of multimedia data.

BACKGROUND

The subject of adaptive transmission of multimedia data over networks has engaged researchers all over the world. During the design and the implementation of an adaptive application special attention must be paid to the following critical modules:

- The module, which is responsible for the transmission of the multimedia data
- The module, which is responsible for monitoring the network conditions and determines the change to the network conditions
- The module, which is responsible for the adaptation of the multimedia data to the network changes
- The module, which is responsible for handling the transmission errors during the transmission of the multimedia data

A common approach for the implementation of adaptive applications is the use of UDP for the transmission of the multimedia data and the use of TCP for the transmission of control information (Parry & Gangatharan, 2005; Vandalore, Feng, Jain, & Fahmy, 1999). Another approach for the transmission of the multimedia data is the use of RTP over UDP (Bouras & Gkamas, 2003; Byers et al., 2000). Most adaptive applications use RTP/RTCP (real time transmission protocol / real time control transmission protocol) (Schulzrinne, Casner, Frederick, & Jacobson, 2003) for the transmission of the multimedia data. The RTP protocol seems to be the de facto standard for the transmission of multimedia data over the Internet and is used both by mbone tools (vit, vat, etc.) and ITU H.323 applications. In addition RTCP offers capabilities for monitoring the transmission quality of multimedia data.

For the implementation of the network monitoring module, a common approach is to use the packet loss as an indication of congestion in the network (Bouras et al., 2003; Byers et al., 2000). One other approach for monitoring the network conditions is the use of utilization of the client buffer (Rejaie, Estrin, & Handley, 1999; Walpole et al., 1997). An important factor that can be used for monitoring the network conditions, and especially for indication of network congestion, is the use of delay jitter during the transmission of the multimedia data.

For the implementation of the adaptation module, some common approaches are the use of rate shaping (Byers et al., 2000; Bouras et al., 2003), the use of layered encoding (Rejaie et al., 1999), the use of frame dropping (Walpole et al., 1997) or a combination of the previous techniques (Ramanujan et al., 1997). The implementation of the adaptation module depends on the encoding method that is used for the transmission of the multimedia data. For example, in order to use the frame dropping technique for the adaptation of a MPEG video stream, a selective frame dropping technique must be used, due to the fact that MPEG video uses inter-frame encoding and some frames contain information relative to other frames. In Vandalore et al. (1999), a detailed survey of application level adaptation techniques is given.

It is important for adaptive real time applications to have "friendly" behavior to the dominant transport protocols (TCP) of the Internet (Floyd & Fall, 1998). In Widmer et al. (2001), a survey

on TCP-friendly congestion control mechanisms is presented.

ADAPTIVE TRANSMISSION OF MULTIMEDIA DATA OVER THE INTERNET

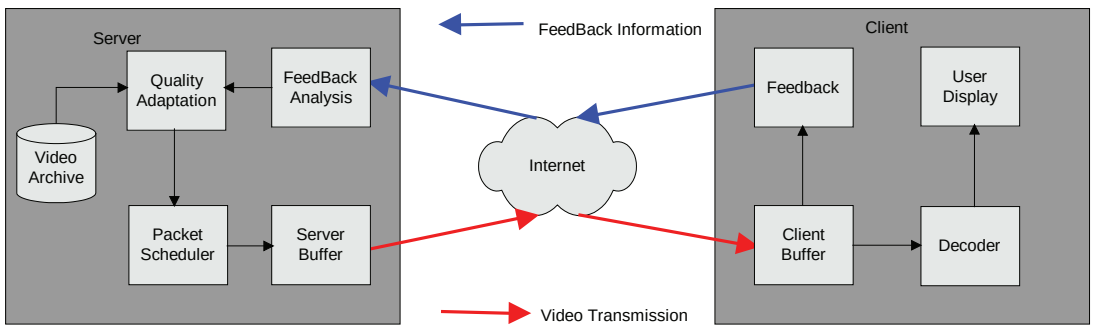
The Architecture of an Adaptive Streaming Application

This section presents a typical architecture for an adaptive streaming application, based on the client server model. Figure 1 displays the architecture of such an adaptive streaming application.

The server of the adaptive streaming architecture consists of the following modules:

- **Video archive:** Video archive consists of a set of hard disks in which the video files are stored. The adaptive streaming application may support various video formats (for example MPEG, JPEG, H.263, etc.). It is possible for one video file to be stored in the video archive in more than one format in order to serve different target user groups. For example, it is possible to store the same video in MPEG format in order to serve the users of the local area network (who have faster network connection with the server) and in H.263 format in order to serve distant users with slow network connections. In this article, we do not investigate the problem of video storage in video archives in order to achieve the optimal performance of the server.
- **Feedback analysis:** This module is responsible for the analysis of feedback information from the network. The role of this module is to determine the network condition mainly based on packet loss rate and delay jitter information, which are provided by RTCP receiver reports. After the examination of network condition, the feedback analysis module informs the quality adaptation module, in order to adapt the transmission of the video to current network conditions.
- **Quality adaptation:** It is responsible for the adaptation of the video transmission quality in order to match with the current network conditions. This module can be implemented using various techniques (rate shaping, layered encoding, frame dropping, etc.).
- **Packet scheduler/Server buffer:** This module is responsible for the encapsulation of multimedia information in the RTP packets. In addition, this module is responsible for the transmission of the RTP packets in the network. In order to smooth accidental problems to the transmission of the multimedia data from the server to the network, an output buffer is used on the server.

Figure 1. System architecture



The client of the adaptive streaming architecture consists of the following modules:

- **Client buffer:** The use of the buffer on the client for the implementation of streaming applications is very important. The client application stores the incoming data to the buffer before starting to present data to the user. The presentation of the multimedia data to the user starts only after the necessary amount of the data is stored in the buffer. The capacity of the client buffer depends to the delay jitter during the transmission of the multimedia data. In any case the capacity of the client buffer must be greater than the maximum delay jitter during the transmission of the data (we suppose that we measure the buffer capacity and the delay jitter in the same units, e.g. in seconds).
- **Feedback:** This module is responsible of monitoring the transmission quality of the data and informing the server. The monitoring of the transmission quality is based on RTCP receiver reports that the client sends to the server. RTCP receiver reports include information about the packet loss rate and the delay jitter during the transmission of the data. With the previous information, the feedback analysis module of the server determines the network's condition.
- **Decoder:** This module reads the data packets from the client buffer and decodes the encoded multimedia information. Depending on the packet losses and the delay during the transmission of the packets, the quality of the multimedia presentation can vary. The decoding and the presentation of the multimedia data can stop, if the appropriate amount of data does not exist in the buffer.
- **User display:** It is responsible for the presentation of the multimedia data to the user.

In the following paragraphs, we give a detailed description of the most important modules of the previously described architecture.

Transmission of Multimedia Data

The transmission of the multimedia data is based on the protocols RTP/RTCP. The protocol RTP is used for the transmission of the multimedia data from the server to the client and the client uses the RTCP protocol, in order to inform the server of the transmission quality.

The RTP/RTCP protocols have been designed for the transmission of real time data like video and audio. Although the RTP/RTCP protocols were initially designed for multicast transmission, they were also used for unicast transmissions. RTP/RTCP can be used for one-way communication like video on demand or for two-way communication like videoconference. RTP/RTCP offers a common platform for the representation of synchronisation information that real time applications needs. The RTCP protocol is the control protocol of RTP. The RTP protocol has been designed to operate in cooperation with the RTCP protocol, which provides information about the transmission quality.

RTP is a protocol that offers end to end transport services with real time characteristics over packet switching networks like IP networks. RTP packet headers include information about the payload type of the data, numbering of the packets and timestamping information.

RTCP offers the following services to applications:

- **QoS monitoring:** This is one of the primary services of RTCP. RTCP provides feedback to applications about the transmission quality. RTCP uses sender reports and receiver reports, which contain useful statistical information like total transmitted packets, packet loss rate and delay jitter during the transmission of the data. This statistical

information is very useful, because it can be used for the implementation of congestion control mechanisms.

- **Source identification:** RTCP source description packets can be used for identification of the participants in a RTP session. In addition, source description packets provide general information about the participants in a RTP session. This service of RTCP is useful for multicast conferences with many members.
- **Inter-media synchronisation:** In real time applications, it is common to transmit audio and video in different data streams. RTCP provides services like timestamping, which can be used for inter-media synchronisation of different data streams (for example synchronisation of audio and video streams).

More information about RTP/RTCP can be found in RFC 3550 (Schulzrinne et al., 2003).

Feedback from the Network

The presentation quality of real time data depends on the packet loss rate and the delay jitter during the transmission over the network. In addition, packet losses or rapid increases of delay jitter may be considered as an indication of problems during the transmission of data over the network. In such a case, the adaptive streaming application must adapt the transmission of the data in order to avoid phenomenon like network congestion. Real time applications have upper bounds to the packet loss rate and to the delay jitter. If packet loss rate or jitter gets to be over these upper bounds, the transmission of real time data can not be continued.

Packet loss rate is defined as the fraction of the total transmitted packets that did not arrive at the receiver. Usually the main reason of packet losses is congestion.

It is difficult to define delay jitter. Some researchers define delay jitter as the difference between the maximum and the minimum delay during the transmission of the packets for a period of time. Some other researchers define delay jitter as the maximum difference between the delay of the transmission of two sequential packets for a period of time. According to RFC 3550 (Schulzrinne et al., 2003), delay jitter is defined to be the mean deviation (smoothed absolute value) of the difference in packet spacing at the receiver compared to the sender for a pair of packets. This is equivalent to the difference in the "relative transit time" for the two packets. The relative transit time is the difference between a packet's timestamp and the receiver's clock at the time of arrival. If s_i is the timestamp from packet i and R_i is the time of arrival for this packet, then for two packets i and j, D is defined as: $D(i,j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i)$. The delay jitter is calculated continuously as each packet arrives, using the difference for that packet and the previous packet, according to the following formula:

$$J_i = J_{i-1} + (|D(i-1, j)| - J_{i-1})/16$$

The previous formula states that the new value of delay jitter depends on the previous value of the delay jitter and on a gain parameter, which gives good noise reduction.

Delay jitter occurs when sequential packets encounter different delays in the queue of the network devices. The different delays are related to the serve model of each queue and the cross traffics in the transmission path.

Sometimes delay jitter occurs during the transmission of real time data, which does not originate from the network but is originated from the transmission host (host included delay jitter). This is because during the encoding of the real time data, the encoder places a timestamp in each packet, which gives information about the time that the packet's information, must be presented to the receiver. In addition, this timestamp is used for the calculation of the delay jitter during the transmission of the real time data. If a notable

time passes from the encoding of the packet and transmission of the packet in the network (because the CPU of the transmitter host is busy) the calculation of the delay jitter is not valid. Host included delay jitter can lead to erroneous estimation for the network conditions.

We can conclude that delay jitter can not lead to reliable estimation of network condition by itself. Delay jitter has to be used in combination with other parameters, like packet loss rate, in order to make reliable estimations of the network conditions. In Bouras et al. (2003), it is shown that the combination of packet loss rate and delay jitter can be used for reliable indication of network congestion.

Quality Adaptation

Quality adaptation module is based on the rate shaping technique. According to the rate shaping technique, if we change some parameters of the encoding procedure, we can control the amount of the data that the video encoder produces (either increase or decrease the amount of the data) and as a result, we can control the transmission rate of the multimedia data.

The implementation of rate shaping techniques depends on the video encoding. Rate shaping techniques change one or more of the following parameters:

- **Frame rate:** Frame rate is the rate of the frames, which are encoded by video encoder. Decreasing the frame rate can reduce the amount of the data that the video encoder produces but will reduce the quality.
- **Quantizer:** The quantizer specifies the number of DCT coefficients that are encoded. Increasing the quantizer decreases the number of encoded coefficients and the image is coarser.
- **Movement detection threshold:** This is used for inter-frame coding, where the DCT is applied to signal differences. The movement detection threshold limits the number of blocks which are detected to be "sufficiently different" from the previous frames. Increasing this threshold decreases the output rate of the encoder.

Error Control/Packet Loss

The packet loss rate is depends on various parameters and the adaptive transmission applications must adapt to changes of packet losses. Two approaches are available to reduce the effects of packet losses:

- **APQ (Automatic Repeat Request):** APQ is an active technique where the receiver and ask the sender to retransmit some lost packets.
- **FEC (Forward Error Correction):** FEC is a passive technique where the sender transmits redundant information. This redundant information is used by the receiver to correct errors and lost packets.

FUTURE TRENDS

The most prominent enhancement of the adaptive real time applications is the use of multicast transmission of the multimedia data. The multicast transmission of multimedia data over the Internet has to accommodate clients with heterogeneous data reception capabilities. To accommodate heterogeneity, the server may transmit one multicast stream and determine the transmission rate that satisfies most of the clients (Byers et al., 2000; Rizzo, 2000; Widmer et al., 2001), and may transmit multiple multicast streams with different transmission rates and allocate clients at each stream or may use layered encoding and transmit each layer to a different multicast stream (Byers et al., 2000). An interesting survey of techniques for multicast multimedia data over the Internet is presented by Li, Ammar, and Paul (1999).

Single multicast stream approaches have the disadvantage that clients with a low bandwidth link will always get a high-bandwidth stream if most of the other members are connected via a high bandwidth link and the same is true the other way around. This problem can be overcome with the use of a multi-stream multicast approach. Single multicast stream approaches have the advantages of easy encoder and decoder implementation and simple protocol operation, due to the fact that during the single multicast stream approach there is no need for synchronization of clients' actions (as is required by the multiple multicast streams and layered encoding approaches).

The subject of adaptive multicast of multimedia data over networks with the use of one multicast stream has engaged researchers all over the world. During the adaptive multicast transmission of multimedia data in a single multicast stream, the server must select the transmission rate that satisfies most the clients with the current network conditions. Three approaches can be found in the literature for the implementation of the adaptation protocol in a single stream multicast mechanism: equation based (Rizzo, 2000; Widmer et al. (2001), network feedback based (Byers et al., 2000), or based on a combination of the previous two approaches (Sisalem & Wolisz, 2000).

CONCLUSION

Many researchers urge that due to the use of new technologies for the implementation of the networks, which offer QoS guarantees, adaptive real time applications will not be used in the future. We believe that this is not true and adaptive real time applications will be used in the future for the following reasons:

- Users may not always want to pay the extra cost for a service with specific QoS guarantees when they have the capability to access a service with good adaptive behaviour.
- Some networks may never be able to provide specific QoS guarantees to the users.
- Even if the Internet eventually supports reservation mechanisms or differentiated services, it is more likely to be on per-class than per-flow basis. Thus, flows are still expected to perform congestion control within their own class.
- With the use of the differential services network model, networks can support services with QoS guarantees together with best effort services and adaptive services.

REFERENCES

Bouras, C., & Gkamas, A. (2003). Multimedia transmission with adaptive QoS based on real time protocols. *International Journal of Communications Systems, Wiley InterScience*, *16*(2), 225-248

Byers, J., Frumin, M., Horn, G., Luby, M., Mitzenmacher, M., Roetter, A., & Shaver, W. (2000). FLID-DL: Congestion control for layered multicast. In *Proceedings of NGC* (pp. 71-81).

Cheung, S. Y., Ammar, M., & Xue, L. (1996). On the use of destination set grouping to improve fariness in multicast video distribution. In *Proceedings of INFOCOM 96*, San Francisco.

Diot, C. (2001, January 25-26). On QoS & traffic engineering and SLS-related work by Sprint. *Workshop on Internet Design for SLS Delivery*, Tulip Inn Tropen, Amsterdam, The Netherlands.

Floyd, S., & Fall, K. (1998, August). Promoting the use of end-to-end congestion control in the Internet. In *IEEE/ACM Transactions on Networking.*

Li, X., Ammar, M. H., & Paul, S. (1999, April). Video multicast over the Internet. *IEEE Network Magazine.*

Parry, M., & Gangatharan, N. (2005). Adaptive data transmission in multimedia networks. *American Journal of Applied Sciences, 2*(3), 730-733.

Ramanujan, R., Newhouse, J., Kaddoura, M., Ahamad, A., Chartier, E., & Thurber, K. (1997). Adaptive streaming of MPEG video over IP networks. In *Proceedings of the 22nd IEEE Conference on Computer Networks,* 398-409.

Rejaie, R., Estrin, D., & Handley, M. (1999). Quality adaptation for congestion controlled video playback over the Internet. In *Proceedings of ACM SIGCOMM '99,* 189-200. Cambridge.

Rizzo, L. (2000) pgmcc: A TCP-friendly single-rate multicast congestion control scheme. In *Proceedings of SIGCOMM 2000,* Stockholm.

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). *RTP: A transport protocol for real-time applications,* RFC 3550, IETF.

Sisalem, D., & Wolisz, A. (2000). LDA+ TCP-friendly adaptation: A measurement and comparison study. The *Tenth International Workshop on Network and Operating Systems Support for Digital Audio and Video,* Chapel Hill, NC.

Vandalore, B., Feng, W., Jain, R., & Fahmy, S., (1999). A survey of application layer techniques for adaptive streaming of multimedia. *Journal of Real Time Systems (Special Issue on Adaptive Multimedia).*

Vickers, B. J., Albuquerque, C. V. N., & Suda, T. (1998). Adaptive multicast of multi-layered video: Rate-based and credit-based approaches. In *Proceedings of IEEE Infocom,* 1073-1083.

Walpole, J., Koster, R., Cen, S., Cowan, C., Maier, D., McNamee, D., et al. (1997). A player for adaptive mpeg video streaming over the Internet. In *Proceedings of the 26th Applied Imagery Pattern Recognition Workshop AIPR-97,* SPIE, (Washington DC), 270-281.

Widmer, J., Denda, R., & Mauve, M., (2001). A survey on TCP-friendly congestion control mechanisms. *Special Issue of the IEEE Network Magazine Control of Best Effort Traffic, 15,* 28-37.

Widmer, J., & Handley, M. (2001). Extending equation-based congestion control to multicast applications. In *Proceedings of the ACM SIGCOMM* (San Diego, CA), 275-285.

KEY TERMS

Adaptive Real Time Applications: Adaptive real time applications are application that have the capability to transmit multimedia data over heterogeneous networks and adapt media transmission to network changes.

Delay Jitter: Delay jitter is defined to be the mean deviation (smoothed absolute value) of the difference in packet spacing at the receiver compared to the sender for a pair of packets.

Frame Rate: Frame rate is the rate of the frames, which are encoded by video encoder.

Movement Detection Threshold: The movement detection threshold is a parameter that limits the number of blocks which are detected to be "sufficiently different" from the previous frames.

Multimedia Data: Multimedia data refers to data that consist of various media types like text, audio, video, and animation.

Packet Loss Rate: Packet loss rate is defined as the fraction of the total transmitted packets that did not arrive at the receiver.

Quality of Service (QoS): Quality of service refers to the capability of a network to provide better service to selected network traffic.

Quantizer: Quantizer specifies the number of DCT coefficients that are encoded.

RTP/RTCP: Protocol which is used for the transmission of multimedia data. The RTP performs the actual transmission and the RTCP is the control and monitoring transmission.

This work was previously published in Encyclopedia of Internet Technologies and Applications, edited by M. Freire and M. Pereira, pp. 16-22, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.13
A Multimedia-Based Threat Management and Information Security Framework

James B. D. Joshi
University of Pittsburgh, USA

Mei-Ling Shyu
University of Miami, USA

Shu-Ching Chen
Florida International University, USA

Walid Aref
Purdue University, USA

Arif Ghafoor
Purdue University, USA

ABSTRACT

This chapter focuses on the key challenges in the design of multimedia-based scalable techniques for threat management and security of information infrastructures. It brings together several multimedia technologies and presents a conceptual architectural framework for an open, secure distributed multimedia application that is composed of multiple domains employing different security and privacy policies and various data analysis and mining tools for extracting sensitive information. The challenge is to integrate such disparate components to enable large-scale multimedia applications and provide a mechanism for threat management. The proposed framework provides a holistic solution for large-scale distributed multi-domain multimedia application environments.

INTRODUCTION

Security of information infrastructures, both in public or private sectors, is vital to overall national security goals. Such infrastructures provide capabilities for gathering, managing, and sharing vital information among numerous organizations that can form large e-enterprises and generally interoperate in the form of a federation of autonomous domains (Joshi, Ghafoor, Aref, & Spafford, 2001; Thuraisingham, 2003). Information shared among multiple domains can come in various forms including text, audio, video, and images which can increase the complexity of security and privacy management. The key security challenges include integration of diverse security policies of collaborating organizations into a coherent capability for protecting information and using collaborative knowledge for detecting and responding to any emerging threats. In addition, information privacy is generally an overriding concern (Adams & Sasse, 1999). Furthermore, a plethora of data analysis and mining tools have emerged that cyber defenders can use to extract sensitive information from public and private multimedia applications and detect patterns and activities indicating potential threats to an infrastructure. Thus, two key challenges to the design of multimedia-based scalable techniques for threat management and security of information infrastructures are *data mining* and *security*, which we briefly overview in the next section.

KEY ISSUES IN DATA MINING AND MULTIMEDIA SECURITY

Multimedia Data Analysis and Mining

Emerging multimedia applications require large-scale integration, mining, and analysis of multimedia data that is generally distributed over multiple security domains. Most of these applications use sensitive information for identifying complex threat actions that cannot be detected via real-time monitoring as such actions can take place over relatively long timeframes. Examples of such applications include detecting the spread of an epidemic and monitoring deterioration of the environment. However, today, data no longer appears in the text form only. Instead, the information from different sources may be in the form of text, image, video, audio, or multimedia documents consisting of several multimedia objects that are tightly synchronized both in space and time (Little & Ghafoor, 1990). Unlike mining the relational data, multimedia data mining is a more complex issue due to the sheer volume and heterogeneous characteristics of the data and the spatial and/or temporal relationships that may exist among multimedia data objects.

Mining multimedia data has recently been addressed in the literature (Chen et al., 2003a; Chen et al., 2004; Thuraisingham, 2003). Most of the existing approaches, however, provide limited capabilities in terms of content analysis and generally do not exploit correlations of multiple data modalities originating from diverse sources and/or sensors. Real-time mining and correlating of multi-modality data from distributed sources and using security-oriented spatio-temporal knowledge can assist in identifying potential threats and ensuring security of large-scale infrastructures (e.g., in command and control environments). In a broader perspective, both *long-ranged* and *real-time* data analysis and mining techniques are needed to allow multi-level content analysis and representation of multimedia data at different levels of resolution to facilitate information classification that has security and privacy implications.

Security Policy and Privacy Management

The multi-modality nature of data and the unique synchronization and *quality of service* (QoS) requirements of multimedia information systems

makes its protection uniquely challenging. These challenges are briefly discussed below.

Content-based Context-aware Access. An application may require controlled access to information based on the sensitivity level of information content, time, location, and other contextual information obtained at the time the access requests are made. For example, in the health care industry, selective content-based access to patient information should be given to physicians, insurance providers, and so forth. Furthermore, a non-primary physician of a patient, who normally does not have access to the patient's records, may need to be allowed access in a *life-threatening* situation.

Heterogeneity of Subjects and Objects. For multimedia applications, heterogeneity implies the diversity of object and subject types. Object heterogeneity implies different types of media, abstract concepts, or knowledge embodied in the information that needs to be protected. For instance, in a digital library system, it is desirable to provide access based on concepts rather than on individual objects. For example, a user may want to access information related to concept *Juvenile Law*. Subject heterogeneity implies that the users have diverse activity profiles, characteristics, and/or qualifications that may not be known *a priori*, which can complicate the specification of access control requirements.

- **Privacy:** Because of the richness of information content of multimedia data, privacy concerns are exacerbated in a multimedia environment (Adams & Sasse, 1999). For instance, a perfectly harmless video sequence of a conference presentation may show a certain behavior of a person (e.g., sleeping during the presentation) in the audience which may result in adversely affecting his/her status if it is viewed by his employer.
- **Dynamically Changing Access/Privacy Requirements:** Access control and privacy requirements in Web-based multimedia applications are inherently dynamic in nature, as the information contents may change frequently. Powerful and flexible access control and privacy models are needed to capture dynamically changing access and privacy parameters in an evolving system.
- **Integration of Access and Privacy Policies:** As multimedia information often exists in distributed heterogeneous administrative domains, there is a growing challenge of developing efficient integration and evolution management tools and techniques for inter-domain access control policies for distributed multimedia applications. Mechanisms are required to support careful analysis of security policies for the multi-domain environment and for mapping multimedia content parameters associated with the *security clusters* that data mining tools may generate to define content-based access control and privacy policies. Situations can also arise when the existing static security policies, both local and global, may not provide sufficient safeguards against emerging threat scenarios and require changes as a scenario unfolds based on the information extracted from newly mined data. In such situations, real-time analysis of dynamic access control policies is required to prevent any potential security breaches in multimedia applications. In such an integrated environment, another key challenge is to efficiently manage the evolution of local and global information classifiers, and security and privacy policies.

Central to these problems is a crucial issue of uniform representation, interchange, sharing, and dissemination of information content over an open, multi-domain environment. Due to the multi-modality nature of multimedia data, the

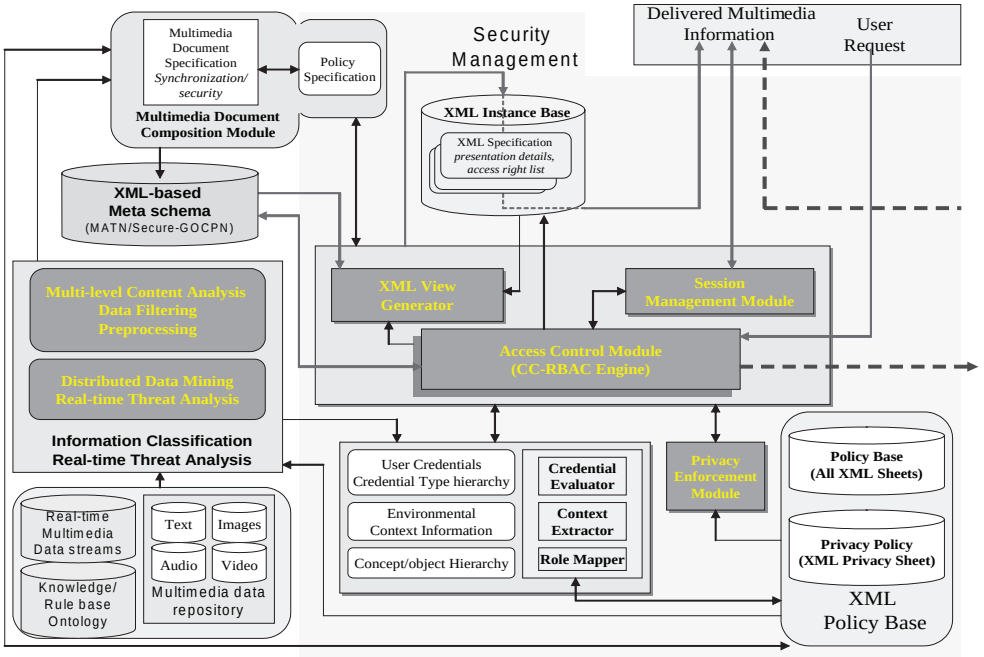
real-time constraints, and heterogeneity of distributed domains, an integrated solution is needed to address the above-mentioned data mining, and security and privacy challenges. To meet these requirements, the integration of several technologies such as data mining, distributed multimedia systems, access control, and privacy models and mechanisms is needed. *EXtensible Markup Language* (XML) (Bertino, Castano, Ferrari, & Mesiti, 1999; Damiani, di Vimercati, Paraboschi, & Samarati, 2002) has emerged as a promising technology for addressing such a fundamental issue of information representation, interchange, sharing, and dissemination (Thuraisingham, Clifton, & Maurer, 2001). XML supports a rich set of features including user-defined tags, nested document structures, and document schemata (Bertino et al., 1999; Bertino, Castano, & Ferrari, 2001; Damiani et al., 2002). Several emerging XML-related technologies, including *Resource Definition Framework* (RDF), *DARPA Agent Markup Language + Ontology Inference Layer* (DAML+OIL), and *Web Ontology Language* (OWL), provide support for integrating information from different domains by facilitating semantic matching of elements (Hunter, 2003; McGuinness, Fikes, Hendler, & Stein, 2002). Moving Pictures Experts Group's MPEG-21 is an open multimedia standard that supports multimedia content delivery and consumption (Burnett et al., 2003). MPEG-21, however, does not capture flexible access control requirements outlined earlier. To the best of our knowledge, no prior research has addressed the above issues in a unified manner by integrating various technologies for developing a secure distributed multi-domain multimedia environment, although several researchers have addressed some of these issues. In this chapter, we discuss various challenges and present a conceptual framework for a scalable secure multi-domain multimedia information system that addresses these challenges.

CONCEPTUAL SYSTEM ARCHITECTURE

Various functional components of the conceptual architecture for a single domain multimedia environment are depicted in Figure 1. The *XML-based Multimedia Document Composition Module* (XDCM) provides an interface for composing XML schemata for multimedia information and policy documents and includes *Policy Composition Interface*. Access control and privacy policy documents are stored in the *XML Policy Base* (XPB). The *Access Control and Privacy Enforcement Modules* (ACM, PEM) constitute the key enforcement modules. The ACM employs a content-based context-aware RBAC engine, extracts the policy information from the XPB, and interacts closely with the *XML View Generator* (XVG) module to produce the authorized presentation schema for a user. Authorized XML instances are further checked against privacy policies by the PEM to ensure that privacy policies are enforced. The *Session Management Module* (SMM) is responsible for monitoring the real-time activities of the multimedia presentation, as well as dynamic access constraints, and interacts with ACM and XVG to update active XML views.

The *Information Classification and Real-time Threat Analysis Module* (ICRTAM) is responsible for the classification of multimedia data and threat analysis based on real-time multimedia data streams. The clusters that it generates are organized in a hierarchy. The *Role Mapper* generates concept roles for these clusters and creates required policy sheets that define access rules and privacy protection requirements. The *Credential Evaluator* (CEv) module evaluates the credentials presented by the ACM. With the help of the *Role Mapper*, it maps the credentials to a role as defined by the access rules. The *Context Extractor* (CEx) evaluates the contextual information and sends the results to the ACM so it can evaluate context-based dynamic policies. The multimedia data repository constitutes the physical objects present in the system that are used to generate the XML multimedia object *meta-*

Figure 1. Conceptual design of a single domain multimedia application environment



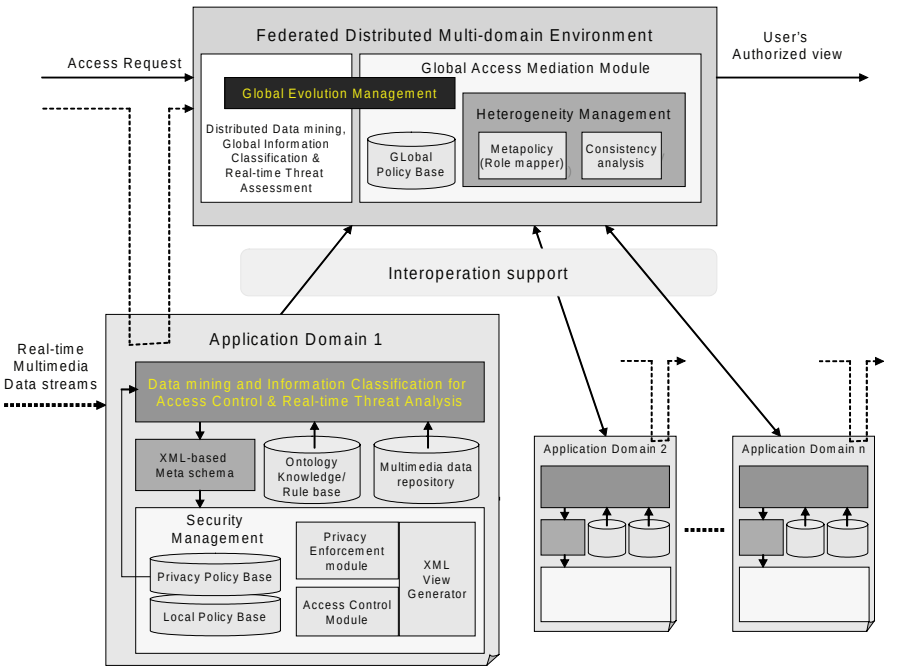
schema, by integrating multimedia presentation information with security and privacy attributes. The *knowledge/rule base* or the *ontology* will be used by the information classification and security cluster generator module.

The *Multi-level Content Analysis, Data Filtering and Pre-processing Module* (MCADFPM) is responsible for (1) conducting multi-level content analysis for mining and learning the semantic information; (2) exploiting suitable multi-resolution representation schemes for various media features, segments, and objects; and (3) enabling automatic feature selection to reach the compact sets of features and dataset for data mining in the data-filtering and preprocessing step. The module assists in generating the security clusters and their attribute and feature sets. The *Distributed Data Mining and Real-time Threat Analysis Module* (DDMRTAM) is responsible for designing and developing (1) distributed data mining techniques for threat assessment based on real-time streams and (2) an information fusion model to resolve the inconsistencies when integrating the classification results from multiple local classifiers. The information clusters that it generates are organized in a hierarchy and represented in an XML.

Conceptual System Architecture for a Multi-Domain Multimedia System

Figure 2 depicts the key components of the conceptual architecture of a secure federated distributed, multi-domain multimedia environment. The *Distributed Data Mining, Global Information Classification and Threat Assessment* component maintains global level views of conceptual objects based on the security clusters of individual domains. It also allows inter-site conceptual mapping of information objects and distributes real-time threat assessment tasks with the help of the ontology or knowledge base of each domain. The *Global Access Mediation Module* integrates all the local policies into a unified *Global Policy Base*. The *Heterogeneity Management* module resolves any semantic differences and incon-

Figure 2. Conceptual design of a multi-domain multimedia application environment



sistencies among the local policies when the global policy is constructed using the *Consistency Analysis* module. The final global level policy generated for the whole multi-domain environment primarily consists of a *Role Mapper* that maintains the mapping from global roles to local roles for access mediation. For a global access request, these mappings are used to issue access requests to individual domains. *Global Evolution Management* module is responsible for overseeing the process of ensuring consistent evolution of local information classification mechanisms and security/privacy policies.

SCALABLE MULTIMEDIA DATA ANALYSIS AND MINING TECHNIQUES

Content-based analysis of multimedia information for facilitating distributed security policies is a challenging issue because of the multi-modality nature of multimedia data and heterogeneity of distributed information sources and sensors. Scalable real-time content-based analysis and mining frameworks are needed to generate security-oriented classification of multimedia information for the distributed multi-domain environment and to provide capabilities for real-time threat assessment. An efficient approach of real-time data stream querying across multiple domains needs to be developed. A set of classifiers, both for individual domains and for global environment, that use a set of training data over relatively long timeframes is essential. Based on the semantics and the contents of information, these classifiers will produce a set of security clusters of multimedia objects, each with a set of features and attributes that can be used for developing information security and privacy policies for single and multi-domain environments, as discussed later. Another key challenge for designing the global classifier for multiple domains is to fuse information from local classifiers and resolve any semantic conflicts and inconsistencies

in an efficient manner. Figure 3 depicts the data analysis and classification processes involved in multimedia information systems.

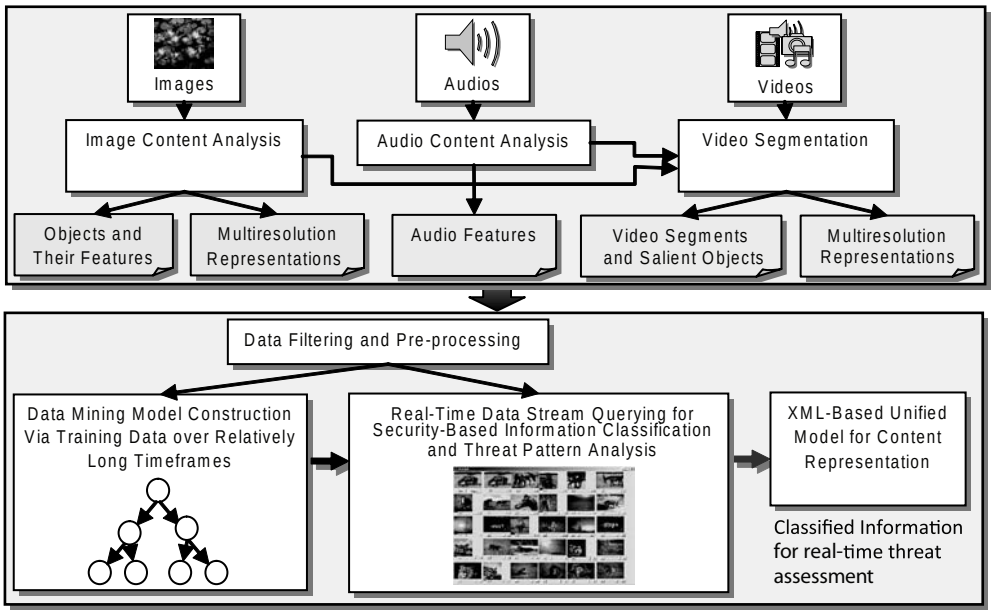
Real-Time Content Analysis of Multi-Modality Multimedia Data. An in-depth content analysis and mining of multimedia data can provide a tremendous wealth of information useful for security management and threat assessment. For this purpose, multi-modality content analysis is essential for integrating and transforming multimedia data into a format suitable for off-line data mining model construction and development of security clusters. Such analysis can span relatively long timeframes. On the other hand, data stream querying for classification and threat assessment requires real-time processing.

Multi-Level Content Analysis of Multi-Modality Data. To automate multi-level real-time content analysis of multimedia data, a framework is needed that allows:

1. extraction of low-level features, including the color and shape features of an image or video frame and the energy feature of an audio segment;
2. extraction of mid-level features such as the object segmentation based on statistical features of the low-level features and the video shot/scene segmentation;
3. extraction of high-level features that identify high-level multimedia content or events; and
4. incremental learning to adapt to the changes in semantics over time.

Mining single-modality alone is not enough to reveal and interpolate the complex correlations of multiple data modalities from diverse sources. Therefore, it is more desirable to discover knowledge from multimedia data with multi-modalities and temporal relationships (Chen et al., 2003; Chen et al., 2004; Dagtas & Abdel-Mottaleb, 2001). A conceptually simple problem in infor-

Figure 3. Architecture of multi-modal content data analysis and real-time data mining



mation fusion is the combination of *opinions* of multiple classifiers. Several information fusion frameworks have been proposed, including the *Bayesian model*, *majority voting rule*, *average of classifier outputs*, *Naïve-Bayes*, *decision template*, and *dynamic classifier selection* (DCS) for integrating information from sources with different modalities (Giacinto, Roli, & Didaci, 2003; Krzysztofowicz & Long, 1990).

Representation of Multimedia Data. Conventional multimedia database systems represent multimedia objects at the single level resolution. Multi-resolution representation consists of maintaining multiple and concurrent representation of multimedia objects and plays an important role in providing semantics to support access control and privacy of information. For example, a confidential medical X-ray image may also have the patient identity information. However, only the version without the identity information should be displayed to unauthorized users. Multi-resolution representation schemes for static images and video data based on the multi-splines method have been proposed in Moni and Kashyap (1995), which has shown to yield excellent visual image quality. In addition, multi-resolution representation provides an attractive solution in terms of coding complexity and image quality.

Real-Time Security-Based Information Classification and Threat Assessment. A set of suitable data mining classifiers, both for individual domains and for the distributed multi-domain environment, are required to support a comprehensive security framework. Such classifiers can be constructed off-line using the training data over relatively long timeframes and used to generate information categorization for implementing the underlying content-based access control policies with a wide range of classification parameters such as sensitivity level of information, its importance, and its semantics. In addition, real-time classifiers are needed to perform threat pattern analysis for detecting any emerging threats. The global classifier will integrate all the local classifiers and resolve the inconsistencies of the information among the local classifiers. These classifiers can generate security clusters consistent with access control and privacy management policies. Steps involved in the information classification process are discussed next.

Data Filtering and Pre-processing. The most challenging step prior to carrying data mining is to clean and prepare the data via data cleaning tools and/or domain-specific security knowledge or rules to ensure that the inputs of the data mining models are consistent with the overall security goals. The lack of a guide to choose the most appropriate subfeature set can severely limit the effectiveness and efficiency of data pre-processing. Moreover, in cases when security knowledge is not available, selecting a subset of compact features can become a bottleneck, especially when the features set is large. Several research works address this issue (Chen et al., 2003; Dy & Brodley, 2000; Shyu et al., 2003). A challenge is to develop efficient data cleaning and prefiltering techniques by enabling automatic security and privacy policy driven feature selection and eliminating unnecessary data records with the aid of domain knowledge.

Real-time Content Correlation of Multiple Streams for Event Detection. There exist no scalable mechanisms that can automatically correlate and synchronize multiple streams in multimodalities and in spatial/temporal dimensions to support real-time data stream querying and event detection. Recently, a technique based on event-template for mining event information by combining audio and video cues to discover high level events in a medical database, such as surgery and patient-physician dialogs in the context of a video database management system was proposed (Aref et al., 2003; Hammad, Franklin, Aref, &

Elmagarmid, 2003; Sandhu, Coyne, Feinstein, & Youman, 1996; Zhu et al., 2003).

Off-line Data Mining Model Construction. Given the training data generated from the data filtering and pre-processing steps, another challenge is to develop a suitable, scalable data mining classifier that can generate relevant security clusters and their feature and attribute sets for implementing access control policies and identifying threat patterns. A data mining classifier is constructed off-line by using the training data over relatively long timeframes. Generally, such classifiers can be used not only in evolving the existing threat patterns but also in discovering new threat patterns, which is extremely vital in real-time threat assessment. Conflicts and inconsistencies may arise while integrating information from multiple classifiers belonging to different domains into a global classifier, requiring information fusion to resolve these problems. A number of fusion rules can be used to combine an individual classification generated by training local classifiers.

XML-based Unified Model for Content Representation. The security clusters identified by the data mining step will consist of complex multimedia objects, which usually come from multiple application domains and in multi-modalities; hence, a unified model for content representation is needed. One promising approach is to use XML for multimedia content representation and develop an XML-based meta-schema for the security clusters and their attribute and feature sets so that these clusters and their sets can be directly used for developing the security management component. In addition, the XML-based meta-schema will represent the multimedia composition schema for secure presentation instances that may be composed of text, audio, video, and images. XML-based meta-schema is also used to represent the information retrieved by the data mining process and to identify multimedia documents and objects corresponding to the features and classes that the classifiers use, forming a high level conceptual "document."

DISTRIBUTED MULTIMEDIA SECURITY AND PRIVACY MANAGEMENT

The development of a fine-grained, content and context based access control and privacy model for multimedia applications raises a number of serious technical challenges. The privacy issue is a major concern in multimedia systems. Efficient techniques are required to generate correct mapping between the security and privacy requirements and the data analysis and classification mechanisms that generate security clusters. A unified XML-based language for expressing the security and privacy policies, as well as complex synchronization constraints, can provide a basis for wider acceptance for practical use.

Access Control and Privacy Protection for Multimedia Information Systems

Access control may need to be applied at different levels of granularity and over hierarchically arranged multimedia objects. Few existing access control models address specific requirements of multimedia information systems. Bertino, Hammad, Aref, and Elmagarmid (2000) proposed a video database access control model that allows access control at various levels of granularity of video information, including video frames, video objects, or a sequence. Joshi, Fahmi, Shafiq, and Ghafoor (2002) proposed a *Secure Generalized Object Composition Petri-net* (Secure-GOCPN) model by extending the GOCPN model of multimedia documents to support multi-level security policies by classifying multimedia information into a prespecified set of classification levels.

Subjects are also assigned security clearances that allow them to view information at or below a particular classification level. The unclassified level may correspond to preventing certain frames from being viewed by an unclassified user or hiding sensitive information from each frame. Using the mapping of classification levels across multiple domains, an implementation architecture that supports the presentation of distributed multimedia documents that include media documents/objects from multiple sources was presented in Joshi et al. (2002). The Secure-GOCPN model, however, only supports multi-level security policy. We propose using a role-based access control (RBAC) approach to address the generic security and privacy requirements of multimedia applications. RBAC supports fine-grained separation of duty (SoD) constraints, principle of least privilege, and efficient permission management. It also facilitates easy integration of policies because of the similarity of roles in different application domains (Ahn & Sandhu, 2000; Ferraiolo et al., 2001; Joshi et al., 2001; Kobsa & Schrek, 2003; Sandhu, Ferraiolo, & Kuhn, 2000). Atluri, Adam, Gomaa, and Adiwijaya (2003) proposed a preliminary formal model of access control of multimedia information by enhancing Synchronized Multimedia Integration Language (SMIL) to express policies.

A Parameterized RBAC Approach to Multimedia Information Systems. In an open environment such as the Web, a system should be able to facilitate a dynamic association of unknown users to a set of authorized actions. A content-based, context-aware RBAC framework would allow the users to be mapped to the activities authorized for them, based on context and credential information they present when requesting accesses. Roles also provide an abstraction level that can be used to capture security relevant features of multimedia objects and the security clusters generated by data mining tools. To handle these issues, one approach is to use parameterized roles and role preconditions that will test the attribute and feature sets of the security clusters and multimedia objects being accessed, the credentials of the requester, and the contextual information that should be valid at the time of access. Along this direction, significant work has been done in the temporal RBAC (TRBAC) model by Bertino, Bonatti, and Ferrari (2001) and later in the generalized TRBAC (GTRBAC) model by Joshi, Bertino, Latif, and Ghafoor (2005). Depending on the application semantics, not all roles may be available to all users at any time. This notion is reflected in the GTRBAC model by defining the following states of a role: *disabled*, *enabled*, and *active*. The *disabled* state indicates the role cannot be activated in a session. The *enabled* state indicates that authorized users activate the role. A role in the *active* state implies that there is at least one user who has activated the role. Accordingly, the preconditions that change the states of a role are: (1) *role enabling/disabling precondition* to enable/disable a role; (2) *role assignment/deassignment precondition* to assign/deassign a user (or permission) to the role; and (3) *role activation/deactivation precondition* that specifies an authorized user can activate a role or when a role should be deactivated.

Though several researchers have addressed access control using parameterized roles, no formal model has been presented by any. One approach is to extend the event-based GTRBAC model to develop a comprehensive event and rule-based, parameterized RBAC framework. A preliminary version of a parameterized role is presented in Joshi, Bhatti, Bertino, and Ghafoor (2004), which essentially builds on the event-based constraint framework of the GTRBAC model.

Extension of an XML Language for CC-RBAC Policies. We propose using an XML language that can express CC-RBAC modeling entities such as user credentials, role events and preconditions, clusters and multimedia objects, and permissions associated with them by adopt-

ing the X-RBAC language proposed by Joshi et al. (2004). A general credential expression of the form (*credTypeID, attributes*), where *credTypeID* is a credential type identifier and *attributes* is a set of attribute-value pairs, is used to map unknown users to predefined roles. In Joshi et al. (2004), an XML language for the basic RBAC model has been proposed to support parameterized RBAC policies. X-RBAC has been developed specifically for protecting XML content—*schema*, *instances*, and *elements*. The *schema*, *instances*, and *elements* of XML documents can naturally be extended to address the requirements of multimedia documents. Based on access control defined for the security clusters of multimedia objects and the presentation schema for such objects, presentation views for the users can be generated. Conceptual level access control policies use concept roles (Bhatti, Joshi, Bertino, & Ghafoor, 2003). It is possible that a schema element does not belong to the cluster to which its parent element belongs. It may further be necessary to specify that the element is not protected by a schema-level policy. At the lowest granularity, the protection may be applied to the fragments of an object such as on portions of a video sequence and individual frames. The XML document structure for the specification of multimedia information also demands a specific set of privileges such as authoring a part of the XML document, modifying attributes only of a particular subset of the XML specification of a multimedia presentation, and navigating privileges to allow the use of links to navigate linked multimedia information.

A considerable amount of work on RBAC can be found in the literature (Ahn & Sandhu, 2000; Ferraiolo et al., 2001; Sandhu et al., 2000). Giuri and Iglio (1997) presented the concepts of parameterized privileges and role templates to allow the specification of policies based on the content of the objects. In Georgiadis, Mavridis, Pangalos, and Thomas (2001), the context of a team is composed of user context and object context. None of these studies addresses the consistency issues and complex requirements of multimedia applications. The environment roles in Covington et al. (2001) are essentially statically defined context roles and thus cannot be employed in a dynamic environment. In Adam, Atluri, and Bertino (2002), a content-based access control (CBAC) model for a digital library has been proposed to protect the textual content of digital archives, uses credentials, and concept hierarchies. Related work also includes several security mechanisms for schema and XML-instance level protection of XML documents (Bertino et al., 1999; Damiani et al., 2002). Qin and Atluri (2003) recently proposed a content-based access control model for Semantic Web and used the concept level protection. These studies, however, do not address the security challenges in emerging multimedia applications mentioned earlier using an RBAC approach.

Privacy Protection in Multimedia Information Systems. In multimedia information systems, privacy concerns exist at multiple layers (Adams & Sasse, 1999). At a primary level, privacy relates to the sensitivity of actual information content of the multi-modal data. At the secondary level, it relates more to the sociopsychological characteristics of the data being accessed (Adams & Sasse, 1999). For example, the clips about medical procedures may be used for teaching purposes, but if the clip shows a patient and his/her employer views it, it may have a detrimental effect, in particular, if the patient has a serious mental disease. While such multimedia clips may be of significant research value, appropriate measures need to be taken to ensure that privacy is protected. The three key factors of privacy protection are: *information sensitivity*, *information receiver*, and *information use* (Adams & Sasse, 1999; Jiang & Landay, 2002). Privacy policies indicate who receives what level of private information for what use. A relevant issue is the inference problem (Thuraisingham, 2003). For instance, disease information may be inferred by collecting information on prescrip-

tion data related to a patient. In particular, data mining can be used to infer facts that are not obvious to human analysis of data. In such cases, we need to protect private information that may be inferred.

A model for user privacy: In general, an application domain provides a privacy policy to protect users' private information that may have been maintained as historical profiles or credential data collected for authorization purposes (Ashley, Hada, Karjoth, & Schunter, 2002; Jiang 2002; Tavani 2001). User consent is needed for releasing such private information to other parties. In particular, privacy requirements may need to be expressed as content-based and association-based constraints (Thuraisingham, 2003). A key challenge is thus to develop a privacy model and an enforcement mechanism that will be driven by the domain rules and user-controlled privacy policies. Suitability of several existing approaches to privacy management has not been applied to multimedia environments. The *sticky policy* based approach proposed in Ashley, Hada, Karjoth, and Schunter (2002) uses privacy statements that *stick* to each piece of private information so that the privacy statements can be checked before the information is accessed. Another relevant approach uses privacy tags and defines an information space boundary (*physical*, *social*, or *activity-based*) as a contextual trigger to enforce permissions defined by the owners of that space (Jiang & Landay, 2002). To address flexible privacy requirements in multimedia applications, these approaches may need to be combined, and the use of parameterized roles can provide significant benefits (Kobsa & Schrek, 2003). One key issue is when a user specifies a privacy policy, the document itself becomes a protected object and may conflict with the overall privacy policy of the domain.

Privacy-aware information classification: To provide a fine-grained control over the release of private information, a mechanism is needed to provide the interaction between the data mining component that generates security clusters and the privacy protection module. Assume that a personal privacy policy states that "*None other than the concerned doctors should view any information that may reveal my disease*". Here, "*any information that may reveal my disease*" may simply be text containing a prescription drug name, the disease name, or a video sequence showing the patient's physical features and indicating signs of the disease; and the "*concerned doctors*" is an abstract concept referring to a user group with certain characteristics. Meta-information from the privacy-policy base can be used as the domain knowledge for supporting the information classifiers at the time of generating security classes and their attributes and feature sets. Furthermore, data mining can reveal new sensitive and private information and associations that may not be possible through human analysis. For such information, if privacy statements do not exist, new privacy policies may need to be generated. Hence, the development of mechanisms that allow efficient interactions between the privacy protection and data mining modules is needed to iteratively refine both privacy policies and information classification. The overall privacy protection component in the proposed research is depicted in Figure 4. The access control module interacts with the privacy mechanism to ensure that privacy requirements are met before users can access data. Several privacy protection models and approaches have been discussed (Adams & Sasse, 1999; Ashley, Hada, Karjoth, & Schunter, 2002; Ashley, Powers, & Schunter, 2002; Jiang & Landay, 2002; Kobsa & Schrek, 2003; Mont, Pearson, & Bramhall, 2003; Tavani & Moor, 2001).

Management of Secure Multi-Domain Multimedia Information Systems

Heterogeneity is a uniquely complex challenge, as security and privacy policies and information classification methodologies may vary over all the

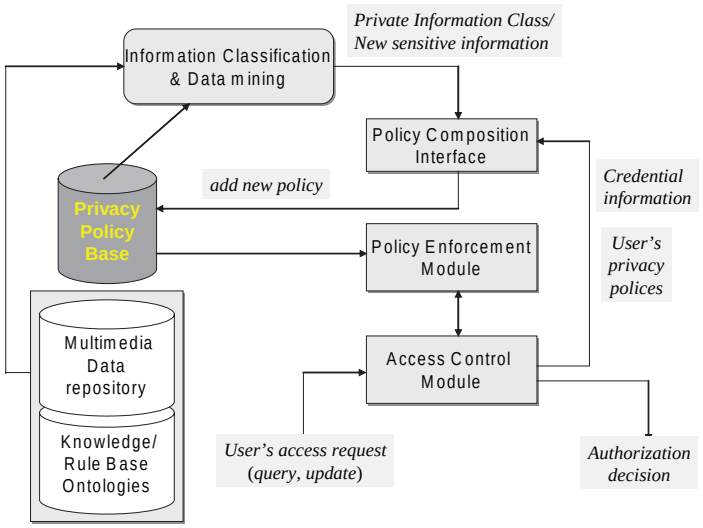
domains giving rise to inconsistent inter-domain actions. The key challenges for developing secure distributed, multi-domain multimedia information systems are (1) to identify conflicts among security and privacy policies and (2) to maintain a consistent global information classification.

The overall infrastructure for a secure distributed multi-domain multimedia application environment must allow seamless and secure interoperation among the diverse access control policies employed by each domain. Such an infrastructure should be scalable, open, and extensible (Joshi et al., 2001) and should allow any domain with an arbitrary security policy and information classification to be a part of the overall environment. An XML-based policy specification and enforcement framework can provide a more practical solution for secure interoperation in multi-domain multimedia environments.

In a multi-domain environment, individual domain policies may follow different conventions and use varying attributes. Semantic heterogeneity may give rise to naming or structural conflicts, a problem that is commonly encountered during schema integration in a heterogeneous federated database system (Batini, Lenzerini, & Navathe, 1986). Naming conflicts occur due to the use of a single name to refer to different concepts. Known techniques from the integration of heterogeneous databases and ontology can be utilized to resolve naming conflicts (Batini et al., 1986). Conflicts can also arise because of the differences in the representation of similar constraints. In the CC-RBAC framework, the constraints can make the semantic heterogeneity issue even more difficult to address. It makes evolution management a significant problem.

Different application domains may use different data analysis and classification mechanisms to generate an equivalent set of concepts with different names or different concepts with similar names. At the global level, such semantic differences need to be properly resolved to ensure secure information access and privacy protection. In each application domain, there is an interaction between the information classification and policy generation modules as indicated in Figure 5. Once policies and information classification consistent to each other have been generated, the global level policy is generated. At the same time, a consistent global level classification also needs to be generated from local classifications. The global

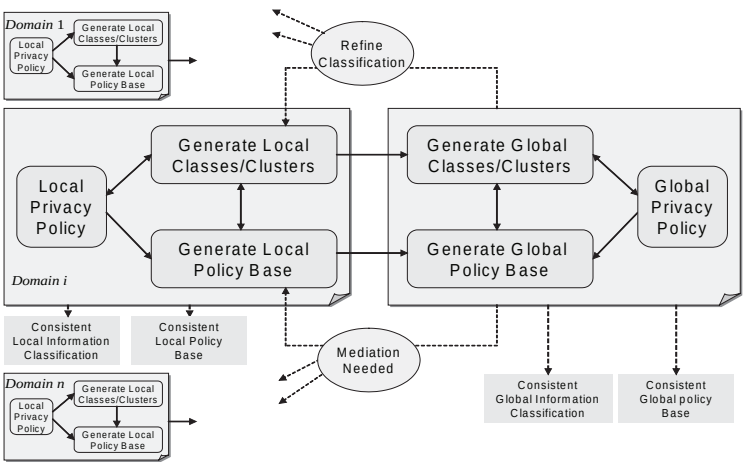
Figure 4. Privacy protection component



level policy and global level classification may be inconsistent if the local domains employ different classification schemes. To resolve such inconsistencies, local classifiers may need to be refined, which in turn affects the local policies. Hence, the global level policy generated earlier will now become inconsistent requiring further mediation of local policies. Such multi-dimensional horizontal and vertical interactions among the classification and policy generation modules at the two levels make the overall integration problem significantly complex. The challenge is to develop a framework for iteratively refining policies and information classes at the two levels. Local clusters/classes are extracted by data mining tools and are used as protection objects. Refinement of classifiers and mediation policies needs to be iterated until consistent classifications and policies are generated at each local domain as well as at the global level. The presence of dynamic and contextual roles can make the problem even more complex. Techniques are needed for detecting inconsistencies, automatically or semi-automatically, using attribute-based matching and ontologies. In particular, use of ontology to assist in semantic mapping between security policies of different domains is becoming a possibility because of tools such as RDF, DAML+OIL, and/or OWL. An ontology can be defined as "the specification of conceptualizations, used to help programs and humans share knowledge" (Hunter, 2003). Use of ontologies in the integration of security policies needs to be similarly explored. In particular, for securing a distributed multi-domain application, use of ontologies related to the domain knowledge as well as the security policies and mechanisms of the individual administrative domains will be crucial to capture the names used to indicate various security attributes and their intended semantics and relationships.

Limited work has been done in the area of secure interoperation in a multi-domain environment. In Shafiq, Joshi, Bertino, and Ghafoor (2004), a policy composition framework for integrating simple RBAC policies to provide secure interoperation has been proposed, where RBAC constraints are transformed into an integer programming problem to optimize secure interoperation. In Joshi et al. (2004), the XML-based RBAC framework has been extended to specify arbitrary multi-domain policies. Integration of multiple policies has been focused generally on MAC mechanisms for federated database systems. Thuraisingham and Ford (1995) discussed methodologies for constraint processing in a multi-level, secure, centralized, and distributed

Figure 5.The integration process



database management system. In Jonscher and Dittrich (1995), the global access layer is used to map global authorizations into local access rights for the individual systems. XML-based tools to manage RBAC policies in an enterprise have been discussed in Vuong, Smith, and Deng (2001).

CONCLUSIONS

In this chapter, we have presented a conceptual architectural framework for open, distributed multimedia information systems that aims at ensuring security and privacy of multimedia information and provides a mechanism for threat assessment. The proposed framework uses XML technologies to address the needs for multi-resolution representation of multi-modal multimedia data, its storage, exchange, and retrieval, as well as to develop access control and privacy specifications. We have discussed the relevance of a content and context based RBAC approach to capture both the access control and privacy requirements of multimedia applications. The framework addresses the need for a holistic solution to supporting flexible access control and privacy requirements on multimedia information. Such information may contain sensitive and private data and can be analyzed to reveal patterns and knowledge for possible threats to an enterprise. The framework provides security and privacy protection while facilitating real-time and long-term analysis to detect potential threats.

REFERENCES

Adam, N.R., Atluri, V., & Bertino, E. (2002). A content-based authorization model for digital libraries. *IEEE Transactions on Knowledge and Data Engineering, 14*(2), 296-315.

Adams, A., & Sasse, M.A. (1999). Taming the wolf in sheep's clothing: Privacy in multimedia communications. *ACM Multimedia, 99,* 101-106.

Ahn, G., & Sandhu, R. (2000). Role-based authorization constraints specification. *ACM Transactions on Information and System Security, 3*(4), 207-226.

Aref, W.G., Catlin, A.C., Elmagarmid, A.K., Fan, J., Hammad, M.A, Ilyas, I., Marzouk, M., Prabhakar, S., & Zhu, X.Q. (2003). VDBMS: A testbed facility for research in video database benchmarking. *ACM Multimedia Systems Journal, Special Issue on Multimedia Document Management Systems.*

Ashley, P., Hada, S., Karjoth, G., & Schunter, M. (2002). E-P3P privacy policies and privacy authorization. *Proceedings of the ACM Workshop on Privacy in the Electronic Society.*

Ashley, P., Powers, C., & Schunter, M. (2002). From privacy promises to privacy management: A new approach for enforcing privacy throughout an enterprise. *Proceedings of the ACM New Security Paradigms Workshop '02.*

Atluri, V., Adam, N., Gomaa, A., & Adiwijaya, I. (2003). Self-manifestation of composite multimedia objects to satisfy security constraints. *ACM Symposium of Applied Computing,* 927-934.

Batini, C., Lenzerini, M., & Navathe, S.B. (1986). A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys, 18*(4), 323-364.

Bertino, E., Bonatti, P.A., & Ferrari, E. (2001). TRBAC: A temporal role-based access control model. *ACM Transactions on Information and System Security, 4*(3), 191-233.

Bertino, E., Castano, S., & Ferrari, E. (2001). Securing XML documents with Author-X. *IEEE Internet Computing,* 21-31.

Bertino, E., Castano, S., Ferrari, E., & Mesiti, M. (1999). Controlled access and dissemination of XML documents. *Proceedings of the 2nd International Workshop on Web Information and Data Management* (pp. 22-27).

Bertino, E., Hammad, M.A., Aref, W.B., & Elmagarmid, A.K. (2000). An access control model for video database systems. *CIKM,* 336-343.

Bhatti, R., Joshi, J.B.D., Bertino, E., & Ghafoor, A. (2003). XML-based RBAC policy specification for secure Web-services. *IEEE Computer.*

Burnett, I., De Walle, R.V., Hill, K., Bormans, J., & Pereira, F. (2003). MPEG-21 goals and achievements. *IEEE Multimedia.*

Chen, S.-C., Shyu, M.-L., Zhang, C., Luo, L., & Chen, M. (2003c). Detection of soccer goal shots using joint multimedia features and classification rules. *Proceedings of the 4th International Workshop on Multimedia Data Mining (MDM/KDD2003),* in conjunction with the *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 36-44).

Chen, S.-C., Shyu, M.-L., Chen, M., & Zhang, C. (2004a). A decision tree-based multimodal data mining framework for soccer goal detection. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2004).*

Covington, M.J., Long, W., Srinivasan, S., Dey, A.K., Ahmad, M., & Abowd, G.D. (2001). Securing context-aware applications using environmental roles. *Proceedings of the 6th ACM Symposium on Access Control Models and Technologies* (pp. 10-20).

Dagtas, S., & Abdel-Mottaleb, M. (2001). Extraction of TV highlights using multimedia features. *Proceedings of the IEEE International Workshop on Multimedia Signal.*

Damiani, E., di Vimercati, S.D.C., Paraboschi, S., & Samarati, P. (2002). A fine-grained access control system for XML documents. *ACM Transactions on Information and System Security, 5*(2), 169-202.

Dy, J.G., & Brodley, C.E. (2000). Feature subset selection and order identification for unsupervised learning. *Proceedings of the 17th International Conference on Machine Learning* (pp. 247-254).

Ferraiolo, D., Sandhu, R., Gavrila, S., Kuhn, R., & Chandramouli, R. (2001). The NIST model for role-based access control: Towards a unified standard. *ACM Transactions on Information and System Security, 4*(3).

Georgiadis, C.K., Mavridis, I., Pangalos, G., & Thomas, R.K. (2001). Flexible team-based access control using contexts. *Proceedings of the 6th ACM Symposium on Access Control Models and Technologies* (pp. 21-27).

Giacinto, G., Roli, F., & Didaci, L. (2003). Fusion of multiple classifiers for intrusion detection in computer networks. *Pattern Recognition Letters, 24,* 1795-1803.

Giuri, L., & Iglio, P. (1997). Role templates for content-based access control. *Proceedings of the 2nd ACM Workshop on RBAC* (pp. 153-159).

Hammad, M.A, Franklin, M., Aref, W.G., & Elmagarmid, A.K. (2003). Scheduling for shared window joins over data streams. *Proceedings of the International Conference on Very Large Data Bases (VLDB)* (pp. 297-308).

Hunter, J. (2003). Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuit and Systems for Video Technology, 13*(1).

Jiang, X, & Landay, J.A. (2002). Modeling privacy control in context-aware systems. *IEEE Pervasive Computing, 1*(3), 59-63.

Jonscher, D., & Dittrich, K.R. (1995). Argos: A configurable access control system for interoperable environments. *Proceedings of the IFIP WG*

11.3 9th Annual Working Conference on Database Security (pp. 39-63).

Joshi, J.B.D., Bertino, E., Latif, U., & Ghafoor, A. (2005). Generalized temporal role based access control model. *IEEE Transactions on Knowledge and Data Engineering, 17*(1), 4-23.

Joshi, J.B.D., Bhatti, R., Bertino, E., & Ghafoor, A. (2004, November-December). An access control language for multidomain environments. *IEEE Internet Computing,* 40-50.

Joshi, J.B.D., Ghafoor, A., Aref, W., & Spafford, E.H. (2001). Digital government security infrastructure design challenges. *IEEE Computer, 34*(2), 66-72.

Joshi, J.B.D., Li, K., Fahmi, H., Shafiq, B., & Ghafoor, A. (2002). A model for secure multimedia document database system in a distributed environment. *IEEE Transactions on Multimedia, Special Issue on Multimedia Databases, 4*(2), 215-234.

Kobsa, A., & Schreck, J. (2003). Privacy through pseudonymity in user-adaptive systems. *ACM Transactions on Internet Technology, 3*(2).

Krzysztofowicz, R., & Long, D. (1990). Fusion of detection probabilities and comparison of multisensor systems. *IEEE Transactions on Systems, Man, & Cybernetics, 20*(3), 665-677.

Little, T.D.C., & Ghafoor, A. (1990). Synchronization and storage models for multimedia objects. *IEEE Journal of Selected Areas in Communications, Special Issue on Multimedia Communication, 8*(3), 413-427.

McGuinness, D.L., Fikes, R., Hendler, J., & Stein, L.A. (2002). DAML+OIL: An ontology language for the Semantic Web. *IEEE Intelligent Systems.*

Moni, S., & Kashyap, R.L. (1995). A multiresolution representation scheme for multimedia databases. *Multimedia Systems, 3,* 228-237.

Mont, M.C., Pearson, S., & Bramhall, P. (2003). Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services. *Proceedings of the 14th International Workshop on Database and Expert Systems Applications* (pp. 377-382).

Qin, L., & Atluri, V. (2003). Concept-level access control for Semantic Web. *Proceedings of the ACM Workshop on XML Security.*

Sandhu, R., Coyne, E.J., Feinstein, H.L., & Youman, C.E. (1996). Role-based access control models. *IEEE Computer, 29*(2), 38-47.

Sandhu, R., Ferraiolo, D., & Kuhn, R. (2000). The NIST model for role-based access control: Toward a unified standard. *Proceedings of the 5th ACM Workshop Role-Based Access Control* (pp. 47-63).

Shafiq, B., Joshi, J.B.D., Bertino, E., & Ghafoor, A. (2004). Secure interoperation in a multi-domain environment employing RBAC policies. *IEEE Transactions on Knowledge and Data Engineering.*

Shyu, M.-L., Chen, S.-C., Chen, M., Zhang, C., & Sarinnapakorn, K. (2003). Image database retrieval utilizing affinity relationships. *Proceedings of the 1st ACM International Workshop on Multimedia Databases (ACM MMDB'03)* (pp. 78-85).

Tavani, H.T., & Moor, J.H. (2001). Privacy protection, control of information, and privacy-enhancing technologies. *ACM SIGCAS Computers and Society, 31*(1).

Thuraisingham, B. (2003). *Web data mining technologies and their applications to business intelligence and counter-terrorism.* Boca Raton, FL: CRC Press.

Thuraisingham, B., Clifton, C., & Maurer, J. (2001). Real-time data mining of multimedia objects. *Proceedings of the 4th International Symposium on Object-Oriented Real-Time Distributed Computing* (pp. 360-365).

Thuraisingham, B., & Ford, W. (1995). Security constraints in a multilevel secure distributed database management system. *IEEE Transactions on Knowledge and Data Engineering*, *7*(2), 274-293.

Vuong, N., Smith, G., & Deng, Y. (2001). Managing security policies in a distributed environment using eXtensible Markup Language (XML). *Proceedings of the 16th ACM Symposium on Applied Computing* (pp. 405-411).

Zhu, X.Q., Fan, J., Aref, W., Catlin, A.C., & Elmagarmid, A. (2003). Medical video mining for efficient database indexing, management and access. *Proceedings of the 19th International Conference on Data Engineering* (pp. 569-580).

This work was previously published in Web and Information Security, edited by E. Ferrari and B. Thuraisingham, pp. 215-241, copyright 2006 by IRM Press (an imprint of IGI Global).

Chapter 2.14
Context-Based Interpretation and Indexing of Video Data

Ankush Mittal
IIT Roorkee, India

Cheong Loong Fah
The National University of Singapore, Singapore

Ashraf Kassim
The National University of Singapore, Singapore

Krishnan V. Pagalthivarthi
IIT Delhi, India

ABSTRACT

Most of the video retrieval systems work with a single shot without considering the temporal context in which the shot appears. However, the meaning of a shot depends on the context in which it is situated and a change in the order of the shots within a scene changes the meaning of the shot. Recently, it has been shown that to find higher-level interpretations of a collection of shots (i.e., a sequence), intershot analysis is at least as important as intrashot analysis. Several such interpretations would be impossible without a context. Contextual characterization of video data involves extracting patterns in the temporal behavior of features of video and mapping these patterns to a high-level interpretation. A Dynamic Bayesian Network (DBN) framework is designed with the temporal context of a segment of a video considered at different granularity depending on the desired application. The novel applications of the system include classifying a group of shots called sequence and parsing a video program into individual segments by building a model of the video program.

INTRODUCTION

Many pattern recognition problems cannot be handled satisfactorily in the absence of contex-

tual information, as the observed values underconstrain the recognition problem leading to ambiguous interpretations. Context is hereby loosely defined as the local domain from which observations are taken, and it often includes spatially or temporally related measurements (Yu & Fu, 1983; Olson & Chun, 2001), though our focus would be on the temporal aspect, that is, measurements and formation of relationships over larger timelines. Note that our definition does not address a contextual meaning arising from culturally determined connotations, such as a rose as a symbol of love.

A landmark in the understanding of film perception was the Kuleshov experiments (Kuleshov, 1974). He showed that the juxtaposition of two unrelated images would force the viewer to find a connection between the two, and the meaning of a shot depends on the context in which it is situated. Experiments concerning contextual details performed by Frith and Robson (1975) showed a film sequence has a structure that can be described through selection rules.

In video data, each shot contains only a small amount of semantic information. A shot is similar to a sentence in a piece of text; it consists of some semantic meaning which may not be comprehensible in the absence of sufficient context. Actions have to be developed sequentially; simultaneous or parallel processes are shown one after the other in a concatenation of shots. Specific domains contain rich temporal transitional structures that help in the classification process. In sports, the events that unfold are governed by the rules of the sport and therefore contain a recurring temporal structure. The rules of production of videos for such applications have also been standardized. For example, in baseball videos, there are only a few recurrent views, such as pitching, close up, home plate, crowd and so forth (Chang & Sundaram, 2000). Similarly, for medical videos, there is a fixed clinical procedure for capturing different video views and thus the temporal structures are exhibited.

The sequential order of events creates a temporal context or structure. Temporal context helps create expectancies about what may come next, and when it will happen. In other words, temporal context may direct attention to important events as they unfold over time.

With the assumption that there is inherent structure in most video classes, especially in a temporal domain, we can design a suitable framework for automatic recognition of video classes. Typically in a Content Based Retrieval (CBR) system, there are several elements which determine the nature of the content and its meaning. The problem can thus be stated as extracting patterns in *the temporal behavior of each variable and also in the dynamics of relationship between the variables, and mapping these patterns to a high-level interpretation*. We tackle the problem in a Dynamic Bayesian Framework that can learn the temporal structure through the fusion of all the features (for tutorial, please refer to Ghahramani (1997)).

The chapter is organized as follows. A brief review of related work is presented first. Next we describe the descriptors that we used in this work to characterize the video. The algorithms for contextual information extraction are then presented along with a strategy for building larger video models. Then we present the overview of the DBN framework and structure of DBN. A discussion on what needs to be learned, and the problems in using a conventional DBN learning approach are also presented in this section. Experiments and results are then presented, followed by discussion and conclusions.

RELATED WORK

Extracting information from the spatial context has found its use in many applications, primarily in remote sensing (Jeon & Landgrebe, 1990; Kittler & Foglein, 1984), character recognition (Kittler & Foglein, n.d.), and detection of faults and cracks

(Bryson et al., 1994). Extracting temporal information is, however, a more complicated task but it has been shown to be important in many applications like discrete monitoring (Nicholson, 1994) and plan recognition tasks, such as tracking football players in a video (Intille & Bobick, 1995) and traffic monitoring (Pynadath & Wellman, 1995). Contextual information extraction has also been studied for problems such as activity recognition using graphical models (Hamid & Huang, 2003), visual intrusion detection (Kettnaker, 2003) and face recognition.

In an interesting analysis done by Nack and Parkes (1997), it is shown how the editing process can be used to automatically generate short sequences of video that realize a particular theme, say humor. Thus for extraction of indices like humor, climax and so forth, context information is very important. A use of context for finding an important shot in a sequence is highlighted by the work of Aigrain and Joly (1996). They detect editing rhythm changes through the second-order regressive modeling of shot duration. The duration of a shot PRED (n) is predicted by (1) where the coefficients of a and b are estimated in a 10-shot sliding window. The rule employed therein is that if (Tn > 2 *PRED (n)) or (Tn < PRED (n)/2) then it is likely that the nth shot is an important (distinguished) shot in a sequence. The above model was solely based on tapping the rhythm information through shot durations. Dorai and Venkatesh (2001) have recently proposed an algorithmic framework called computational media aesthetics for understanding of the dynamic structure of the narrative structure via analysis of the integration and sequencing of audio/video elements. They consider expressive elements such as tempo, rhythm and tone. Tempo or pace is the rate of performance or delivery and it is a reflection of the speed and time of the underlying events being portrayed and affects the overall sense of time of a movie. They define P (n), a continuous valued pace function, a

$$P(n) = W(s(n)) + \frac{m(n) - \mu_m}{\sigma_m}$$

where s refers to shot length in frames, m to motion magnitude, μ_m and σ_m, to the mean and standard deviation of motion respectively and n to shot number. W (s(n)) is an overall two-part shot length normalizing scheme, having the property of being more sensitive near the median shot length, but slows in gradient as shot length increases into the "longer" range.

We would like to view the contextual information extraction from a more generic perspective and consider the extraction of temporal pattern in the behavior of the variables.

THE DESCRIPTORS

The perceptual level features considered in this work are Time-to-Collision (TTC), shot editing, and temporal motion activity. Since our emphasis is on presenting algorithms for contextual information, they are only briefly discussed here. The interested reader can refer to Mittal and Cheong (2001) and Mittal and Altman (2003) for more details. TTC is the time needed for the observer to reach the object, if the instantaneous relative velocity along the optical axis is kept unchanged (Meyer, 1994). There exists a specific mechanism in the human visual system, designed to cause one to blink or to avoid a looming object approaching too quickly. Video shot with a small TTC evokes fear because it indicates a scene of impending collision. Thus, TTC can serve as a potent cue for the characterization of an accident or violence.

Although complex editing cues are employed by the cameramen for making a coherent sequence, the two most significant ones are the shot transitions and shot pacing. The use of a particular shot transition, like dissolve, wipe, fade-in, and so forth, can be associated with the possible intentions of the movie producer. For example,

dissolves have been typically used to bring about smoothness in the passage of time or place. A shot depicting a close-up of a young woman followed by a dissolve to a shot containing an old woman suggests that the young woman has become old. Similarly, shot pacing can be adjusted accordingly for creating the desired effects, like building up of the tension by using a fast cutting-rate. The third feature, that is, temporal motion feature, characterizes motion via several measures such as total motion activity, distribution of motion, local-motion/global-motion, and so forth.

Figure 1 shows the effectiveness of TTC in the content characterization with the example of a climax. Before the climax of a movie, there are generally some chase scenes leading to the meeting of bad protagonists of a movie with good ones. One example of such a movie is depicted in this figure, where the camera is shown both from the perspective of the prey and of the predator leading to several large lengths of the impending collisions as depicted in Figure 2. During the climax, there is a direct contact, and therefore collision length is small and frequent. Combined with large motion and small shot length (both relative to the context), TTC could be used to extract the climax sequences.

Many works in the past have focused on shot classification based on single-shot features like color, intensity variation, and so forth. We believe that since each video class represents structured events unfolding in time, the appropriate class signatures are also present in the temporal domain. The features such as shot-length, motion, TTC, and so forth, were chosen to illustrate the importance of temporal information, as they are perceptual in nature, as opposed to low-level features such as color, and so forth.

CONTEXT INFORMATION EXTRACTION

Hierarchy of Context

Depending on the level of abstraction of the descriptors, the context information can be coarsely

Figure 1. Shots leading to climax in a movie

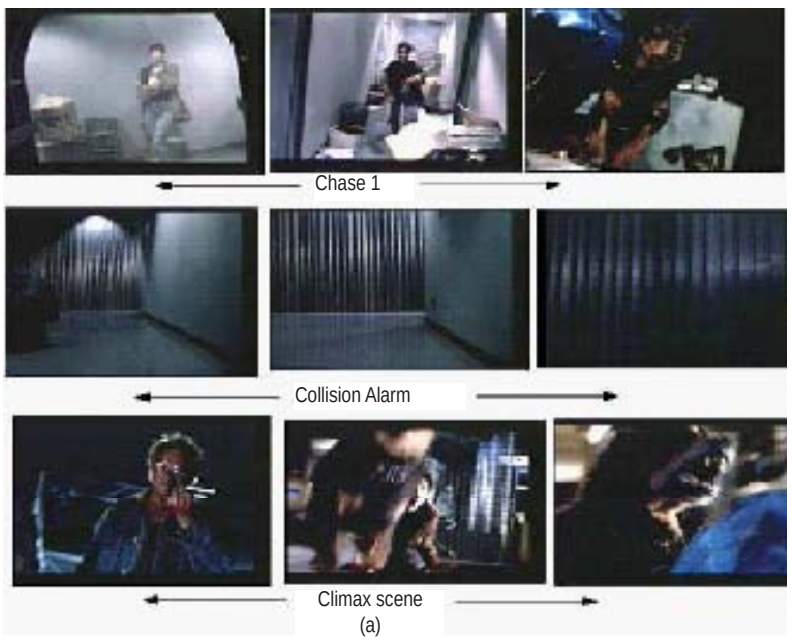
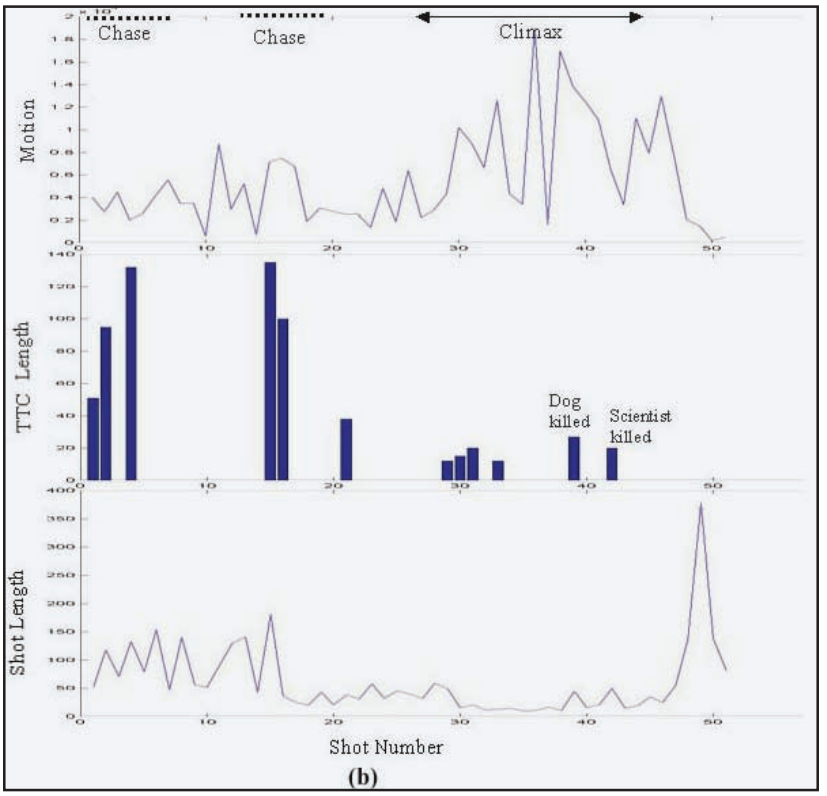


Figure 2. Feature values for the shots in Figure 1. Shots in which TTC length is not shown correspond to a case when TTC is infinite; that is, there is no possibility of collision.



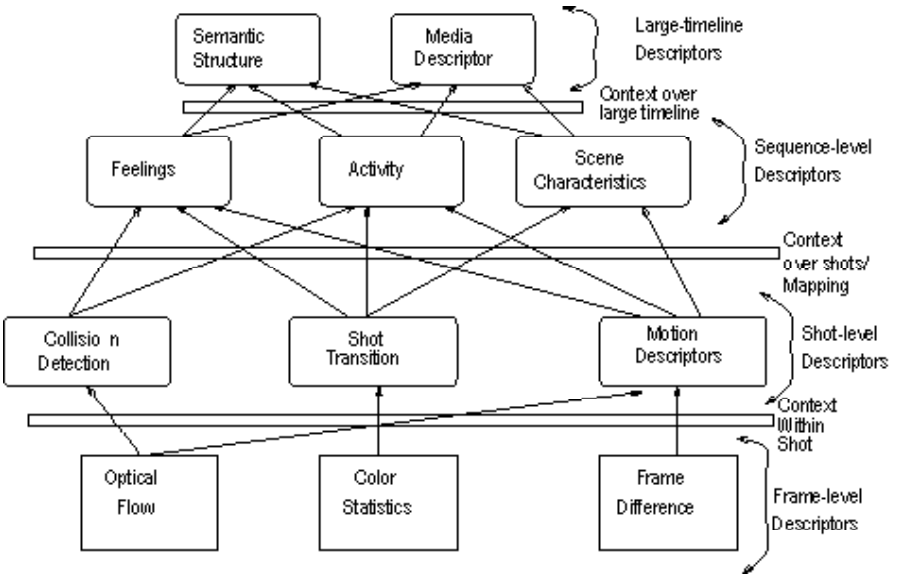
(i.e., large neighborhood) or finely integrated. Figure 3 shows a hierarchy of descriptors in a bottom-up fashion where each representational level derives its properties from the lower levels. At the lowest level of the hierarchy are properties like color, optical flow, and so forth, which correspond only to the individual frames. They might employ the spatial context methods that are found in the image processing literature.

At the second lowest level are descriptors such as TTC, shot transition details, and so forth, which derive their characteristics from a number of frames. For example, patterns in the mean and variance of color features computed over a number of frames (i.e., context) are used in identifying shot-transition (Yeo & Liu, 1995). Similarly, the context over frames is used for collision detection by identifying a monotonically decreasing TTC, and for the extraction of motion descriptors (like local-motion/global-motion).

At the next higher level are the sequence level descriptors (or indices), which might require the information of several shots. Some examples of these descriptors could be in terms of feelings (like interesting, horror, excitement, sad, etc.), activity (like accident, chase, etc.) and scene characteristics (like climax, violence, newscaster, commercial, etc.). For example, a large number of collisions typically depict a violent scene, whereas one or two collision shots followed by a fading to a long shot typically depicts a melancholy scene. If the context information over the neighboring shots is not considered, then these and many other distinctions would not be possible.

Finally, these indices are integrated over a large timeline to obtain semantic structures (like

Figure 3. A hierarchy of descriptors



for news and sports) or media descriptors. An action movie has many scenes where violence is shown, a thriller has many climax scenes and an emotional movie has many sentimental scenes (with close-ups, special effects, etc.). These make it possible to perform automatic labeling as well as efficient screening of the media (for example, for violence or for profanity).

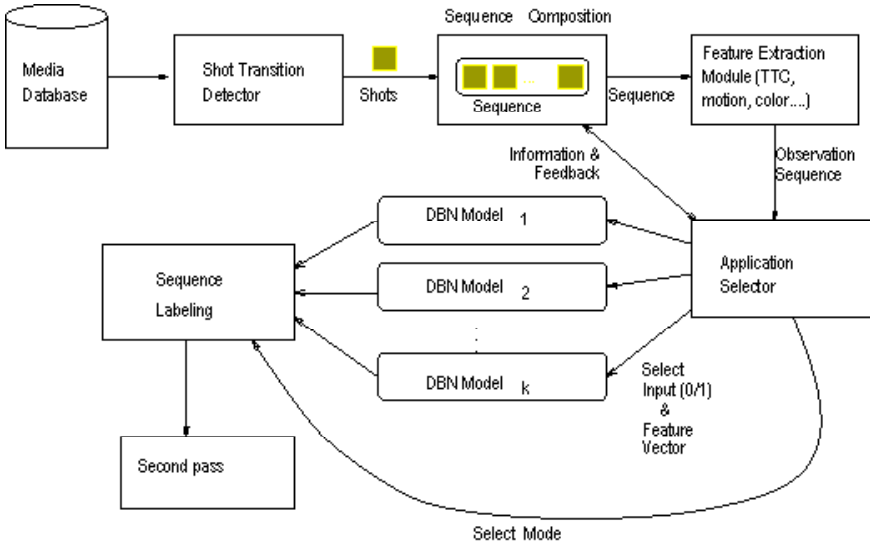
Descriptors at different levels of the hierarchy need to be handled with different strategies. One basic framework is presented in the following sections.

The Algorithm

Figure 4 depicts the steps in a context information extraction (mainly the first pass over the shots). The digitized media is kept in the database and the shot transition module segments the raw video into shots. The sequence composer groups a number of shots (say /) based on the similarity values of the shots (Jain, Vailaya, & Wei, 1999) depending on the selected application. For example, for applications like climax detection, the sequences consist of a much larger number of shots than that for the sports identifier.

The feature extraction of a sequence yields an observation sequence, which consists of feature vectors corresponding to each shot in the sequence. An appropriate DBN Model is selected based on the application, which determines the complexity of the mapping required. Thus, each application has its own model (although the DBN could be made in such a way that a few applications could share a common model, but the performance would not be optimal). During the training phase, only the sequences corresponding to positive examples of the domain are used for learning. During the querying or labeling phase, DBN evaluates the likelihood of the input observation sequence belonging to the domain it represents. The sequence labeling module and the application selector communicate with each other to define the task which needs to be performed, with the output being a set of likelihoods. If the application involves classifying into one of the exclusive genres, the label corresponding to the DBN model with the maximum likelihood is as-

Figure 4. Steps in context information extraction (details of the first pass are shown here)



signed. In general, however, a threshold is chosen (automatically during the training phase) over the likelihoods of a DBN model. If the likelihood is more than the threshold for a DBN model, the corresponding label is assigned. Thus, a sequence can have zero label or multiple labels (such as "interesting", "soccer", and "violent") after the first pass. Domain rules aid further classification in the subsequent passes, the details of which are presented in the next section.

Context Over Sequences: The Subsequent Passes

The algorithm in the subsequent passes is dependent on the application (or the goal). The fundamental philosophy, however, is to map the sequences with their labels onto a one-dimensional domain, and apply the algorithms relevant in the spatial context. A similar example could be found in the character recognition task, where the classification accuracy of a character could be improved by looking at the letters both preceding and following it (Kittler & Foglein, 1984). In this case, common arrangements of the letters (e.g., in English: qu, ee, tion) are used to provide a context within which the character may be interpreted.

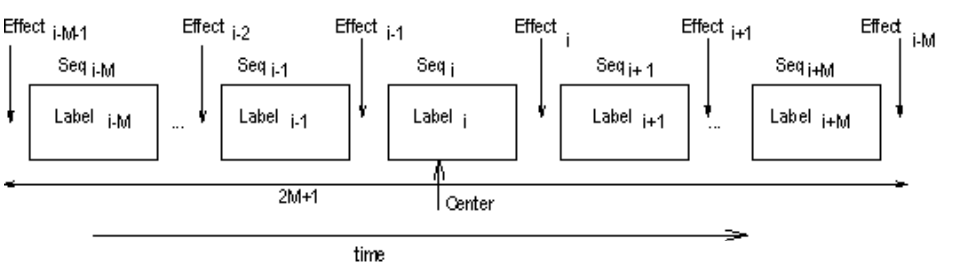
Three examples of algorithms at this level are briefly presented in this section as follows:

1. Probabilistic relaxation (Hummel & Zucker, 1983) is used to characterize a domain like the identification of climax, which has a lot of variation amongst the training samples. In probabilistic relaxation, each shot within the contextual neighborhood is assigned a label with a given probability (or likelihood) by the DBN models. A sliding window is chosen around the center shot. Iterations of the relaxation process update each of the labels around the central shot with respect to a compatibility function between the labels in the contextual neighborhood. In this manner, successive iterations propagate context throughout the timeline. The compatibility function encodes the constraint relationships, such as it is more likely to find a climax scene at a "small" distance from the other climax scene. In other words, the probabilities of both climax scenes are

increased after each pass. The relaxation process is stopped when the number of passes exceeds a fixed value or when the passes do not bring about significant changes to the probabilities.

2. Context cues can also aid in improving the classification accuracy of shots. Consider a CBR system that considers an individual shot or a group of shots (say, 4 to 10), that is, sequence, seq_i. A sequence seq_i is classified into one of the classes (like MTV, soccer, etc.) with $label_i$ on the basis of its low-level and high-level properties. Figure 5 shows the time-layout of sequences with their labels. $Effect_i$ is the transition-effect between seq_i and seq_{i+1}. Consider the 2M neighborhood sequences (in one dimension) of seq_i. The nature of the video domain puts a constraint on defining the constraint relationships within the neighborhood. For example, it is generally not possible to find a tennis shot in between several soccer shots. On the other hand, a few commercial shots may certainly be present in a soccer match. Our strategy to reduce the misclassifications is to slide a window of 2M+1 size over each sequence and check against governing the labels and effects in the neighborhood. If $label_i$ and $label_{i-1}$ are to be different (i.e., there is a change of video class), $effect_i$ should be a special effect (and not a flat-cut), and $label_i$ should match at least a few labels ahead, that is, $label_{i+1}$ and so on. In the second pass, those consecutive items with the same labels are clustered together. A reallotment of the labels is done on the basis of rules, like the length of each clustered sequence should be greater than a minimum value. The second pass can also be used to combine unclassified clustered sequences with appropriate neighboring items. This strategy can also be used to retrieve the entire program. By classifying only the individual shots, retrieval of the entire program is a difficult task. The model can only be built for certain sections of the program. For example, it is not an easy task to model outdoor shots in News, or in-between audience shots in sports. However, through this pass scheme, parts of the same program can also be appropriately combined to form one unit.
3. The goal of the subsequent pass could be to extract information about the media (such as the media being an action movie, thriller movie or an emotional movie), or to construct semantic timelines. For media descriptors like action, the pass involves counting the number of violent scenes, with due consideration to their degrees of violence, which is estimated by the likelihood generated by the DBN model. The pass also considers motion, shot length, and so forth, in the scenes classified as "nonviolent." Constraint conditions are also enhanced during the pass (for instance, the shot length on average is smaller in an action movie than that in a

Figure 5. Subsequent passes perform analysis on the context line of sequences



thriller). The application of constructing semantic timelines is considered in detail in the next section.

Building Larger Models

How DBNs can be made to learn larger models of video is illustrated here. Consider, for example, parsing broadcast news into different sections, and the user is provided with a facility to browse any one of them. Examples of such queries could be "Show me the sport clip which came in the news" and "go to the weather report."

If a video can be segmented into its scene units, the user can more conveniently browse through that video on a scene basis rather than on a shot-by-shot basis, as is commonly done in practice. This allows a significant reduction of information to be conveyed or presented to the user.

Zweig (1998) has shown that a Finite State Automaton (FSA) can be modeled using DBNs. The same idea is explored in the domain of a CBR system. The video programs evolve through a series of distinct processes, each of which is best represented by a separate model. When modeling these processes, it is convenient to create submodels for each stage, and to model the entire process as a composition of atomic parts. By factoring a complex model into a combination of simpler ones, we achieve a combinatorial reduction in the number of models that need to be learned. Thus a probabilistic nondeterministic FSA can be constructed as shown in Mittal and Altman (2003). In the FSA, each of the states represents a stage of development and can be either manually or automatically constructed with a few hours of News program.

Since most video programs begin and end with a specific video sequence (which can be recognized), modeling the entire structure through FSA, which has explicitly defined start and end states, is justified. The probabilistic FSA of News has transitions on the type of shot cut (i.e., dissolve, wipe, etc.) with a probability.

The modeling of the FSA by DBN can be done as follows. The position in the FSA at a specific time is represented by a state variable in the DBN. The DBN transition variable encodes which arc is taken out of the FSA state at any particular time. The number of values the transition variable assumes is equal to the maximum out-degree of any of the states in the FSA. The transition probabilities associated with the arcs in the automaton are reflected in the class probability tables associated with the transition variables in the DBN.

DYNAMIC BAYESIAN NETWORKS

DBNs (Ghahramani, 1997; Nicholson, 1994; Pavlovic, Frey, & Huang, 1999) are a class of graphical, probabilistic models that encode dependencies among sets of random variables evolving in time. They generalize the Hidden Markov Models (HMM) and the Linear dynamical systems by adopting a wider range of topologies and inference algorithms.

DBN has been used to temporally fuse heterogeneous features like face, skin and silence detectors for tackling the problem of speaker detection (Garg et al., 2000; Pavlovic et al., 2000). An important application which demonstrates the potential of DBN is multivariate classification of business cycles in phases (Sondhauss & Weihs, 1999). Modeling a DBN with enough hidden states allows the learning of the patterns of variation shown by each feature in individual video classes (or high-level indices). It can also establish correlations and associations between the features leading to the learning of conditional relationships. Temporal contextual information from temporal neighbors is conveyed to the current classification process via the class transition probabilities. Besides the fact that DBNs were explicitly developed to model temporal domain, the other reason for preferring DBN over time-delay neural networks or modified versions of other learning tools like SVM, and so forth, is

that DBN offers the interpretation in terms of probability that makes it suitable to be part of a larger model.

Structuring the CBR Network

Consider an observation sequence seq_T consisting of feature vectors $F^0, \ldots, F^T$ for $T + 1$ shots (typically seven to 30 shots). Since multiple-label assignment should be allowed in the domain of multimedia indexing (for example, "interesting" + "soccer"), each video class or index is represented by a DBN model. A DBN model of a class is trained with preclassified sequences to extract the characteristic patterns in the features. During the inference phase, each DBN model gives a likelihood measure, and if this exceeds the threshold for the model, the label is assigned to the sequence.

Let the set of n CBR features at time t be represented by $F_1^t, \ldots, F_n^t$ where feature vector $\mathbf{F} \in \mathbf{Z}$, Z is the set of all the observed and hidden variables. The system is modeled as evolving in discrete time steps and is a compact representation for the two time-slice conditional probability distribution $P(Z^{t+1} | Z^t)$. Both the state evolution model and the observation model form a part of the system such that the Markov assumption and the time-invariant assumption hold. The Markov assumption simply states that the future is independent of the past given the present. The time-invariant assumption means that the process is stationary, that is, $P(Z^{t+1} | Z^t)$ is the same for all t, which simplifies the learning process.

A DBN can be expressed in terms of two Bayesian Networks (BN): a prior network BN0, which specifies a distribution over the initial states, and a transition network BN, which represents the transition probability from state Z^t to state Z^{t+1}. Although a DBN defines a distribution over infinite trajectories of states, in practice, reasoning is carried out on a finite time interval 0, . . . , T by "unrolling" the DBN structure into long sequences of BNs over $Z^1, \ldots, Z^T$. In time slice 0, the parents of Z^0 and its conditional probability distributions (CPD) are those specified in the prior network BN_0. In time slice t + 1, the parents of Z^{t+1} and its CPDs are specified in BN_t. Thus, the joint distribution over $Z^1, \ldots, Z$ is

$$P(z^0, \ldots, z^T) = P_{BN_0}(z^0) \prod_{t=1}^{T-1} P_{BN_\rightarrow}(z^{t+1} | z^t)$$

Consider a DBN model and an observation sequence seq_T consisting of feature vectors $F^1, \ldots, F^m$. There are three basic problems of inference, learning and decoding of a model which are useful in a CBR task.

DBN Computational Tasks

The inference problem can be stated as computing the probability $P(\ddot{e} | seq_i)$ such that the observation sequence is produced by the model. The classical solution to the DBN inference is based on the same theory as the forward-backward propagation for HMMs (Rabiner, 1989). The algorithm propagates forward messages at the start of the sequence, gathering evidence along the way. Similar process is used to propagate the backward messages β_t in the reverse direction. The posterior distribution $\ddot{O}^t$ over the states at time t is simply $at(z) \cdot \beta^t(z)$ (with suitable re-normalization). The joint posterior over the states at t and t + 1 is proportional to $\acute{a}_t(z^t) \cdot P_{BN_\rightarrow}(z^{t+1} | z^t) \cdot \beta^{t+1}(z^{t+1}) \cdot P_{BN_\rightarrow}(F^{t+1} | z^{t+1})$.

The learning algorithm for dynamic Bayesian networks follows from the EM algorithm. The goal of sequence decoding in DBN is to find the most likely state sequence Q of hidden variables given the observations such that $X_{*T} = arg \max_{XT} P(X_T | seq_T)$. This task can be achieved by using the Viterbi algorithm (Viterbi, 1967) based on dynamic programming. Decoding attempts to uncover the hidden part of the model and outputs the state sequence that best explains the observations. The previous section presents an application to parse a video program where each state corresponds to a segment of the program.

What Can Be Learned?

An important question that can be raised is this: Which pattern can be learned and which cannot be learnt? Extracting temporal information is complicated. Some of the aspects of the temporal information that we attempt to model are

1. Temporal ordering. Many scene events have precedent-antecedent relationships. For example, as discussed in the sports domain, shots depicting interesting events are generally followed by a replay. These replays can be detected in a domain-independent manner and the shots preceding them can be retrieved, which would typically be the highlights of the sport.
2. Logical constraints could be of the form that event A occurs either before the event B or the event C or later than the event D.
3. Time duration. The time an observation sequence lasts is also learnt. For example, a climax scene is not expected to last over hundreds of shots. Thus, a very-long sequence having feature values (like large number of TTC shots and small shot-lengths) similar to the climax should be classified as nonclimax.
4. Association between variables. For instance, during the building up of a climax, TTC measure should continuously decrease while the motion should increase.

The Problems in Learning

Although DBN can simultaneously work with continuous variables and discrete variables, some characteristic effects, which do not occur often, might not be properly modeled by DBN learning. A case in point is the presence of a black frame at the end of every commercial. Since it is just a spike in a time-frame, DBN associates more meaning to its absence and treats the presence at the end of sequence as "noise". Therefore, we included a binary variable: "presence of a black frame," which takes the value 1 for every frame if at the end of a sequence there is a black frame.

While DBN discovers the temporal correlations of the variables very well, it does not support the modeling of long-range dependencies and aggregate influences from variables evolving at different speeds, which could be a frequent occurrence in a video domain. This solution to the problem is suggested as explicitly searching for violations of the Markov property at widely varying time granularity (Boyen et al., 1999).

Another crucial point to be considered in the design of DBN is the number of hidden states. Insufficient numbers of hidden states would not model all the variables, while an excessive number would generally lead to overlearning. Since, for each video class a separate DBN is employed, the optimization of the number of hidden states is done by trial and test method so as to achieve the best performance on the training data. Since feature extraction on video classes is very inefficient (three frames/second!), we do not have enough data to learn the structure of DBN at present for our application. We assumed a simple DBN structure as shown in Figure 6.

Another characteristic of the DBN learning is that the features which are not relevant to the class have their estimated probability density functions spread out, which can easily be noticed and these features can be removed. For example, there is no significance of the shot-transition length in cricket sequences, and thus this feature can be removed.

Like most of the learning tools, DBN also requires the presence of a few representational training sequences of the model to help extract the rules. For example, just by training the interesting sequences of cricket, the interesting sequences of badminton cannot be extracted. This is because the parameter learning is not generalized enough. DBN is specific in learning the shot-length specific to cricket, along with the length of the special effect (i.e., wipe) used to indicate replay. On the

other hand, a human expert would probably be in a position to generalize these rules.

Of course, this learning strategy of DBN has its own advantages. Consider for example, a sequence of cricket shots (which is noninteresting) having two wipe transitions and the transitions are separated by many flat-cut shots. Since DBN is trained to expect only one to three shots in between two-wipe transitions, it would not recognize this as an interesting shot.

Another practical problem that we faced was to decide the optimum length of shots for DBN learning. If some redundant shots are taken, DBN fails to learn the pattern. For example, in climax application, if we train the DBN only with climax scene, it can perform the classification with good accuracy. However, if we increase the number of shots, the classification accuracy drops. This implies that the training samples should have just enough number of shots to model or characterize the pattern (which at present requires human input).

Modeling video programs is usually tough, as they cannot be mapped to the templates except in exceptional cases. The number of factors which contribute toward variations in the parameters and thus stochastic learning in DBN is highly suitable. We would therefore consider subsequent passes in the next section which takes the initial probability assigned by DBN and tries to improve the classification performance based on the task at hand.

EXPERIMENTS AND APPLICATIONS

For the purpose of experimentation, we recorded sequences from TV using a VCR and grabbed shots in MPEG format. The size of the database was around four hours 30 minutes (details are given in Table 1) from video sequences of different categories. The frame dimension was 352×288.

The principal objective of these experiments is not to demonstrate the computational feasibility of some of the algorithms (for example, TTC or shot detection). Rather we want to demonstrate that the constraints/laws derived from the structure or patterns of interaction between producers and viewers are valid. A few applications are considered to demonstrate the working and effectiveness of the present work.

In order to highlight the contribution of the perceptual-level features and the contextual information discussed in this chapter, other features like color, texture, shape, and so forth, are not employed. The set of features employed is shown in Figure 6. Many works in the past (including ours (Mittal & Cheong, 2003)) have focused on shot classification based on single-shot features like color, intensity variation, and so forth. We believe

Table 1. Video database used for experimentation

Type	Duration (min.' sec")
Sports	63 '50"
MTV	12 '10"
Commercial	17 '13"
Movie	27 '25"
BBC News	64 '54"
News from Singapore TCS Channel 5 ((consist of 2 commercial breaks)	87 '50"
Total	**4 hr 33 min 22 sec**

Figure 6. DBN architecture for the CBR system. The black frame feature has value 1 if the shot ends with black frame and 0 otherwise. It is especially relevant in detection of commercials.

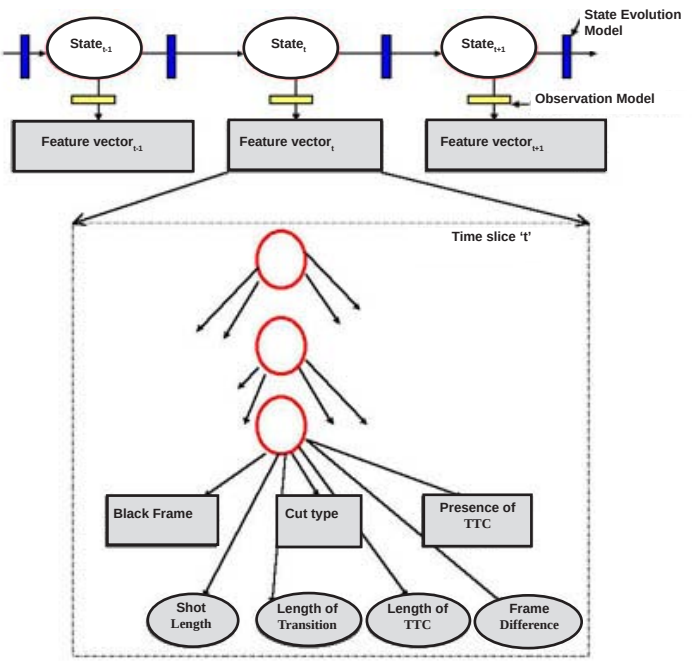


Table 2. Performance: Sequence classifier

Class	Fixed Number of shots		Fixed Number of frames	
	Misclassified (Out of 30)	False alarm (Out of 10 seq. for each class)	Misclassified (Out of 30)	False alarm (Out of 10 seq. for each class)
Commercial (CO)	None	None	3	None
Cricket (CR)	4	3 MT, 1 SO	8	1 MT, 1 NE, 7 SO, 4 TE
MTV (MT)	8	2 NE, 2 TE	11	7 NE, 6 TE, 7 SO
News (NE)	3	None	9	3 TE, 1 SO
Soccer (SO)	10	2 CO, 4 CR, 5 MT, 8 NE, 7 TN	13	1 CO, 1 MT, 4 NE, 9 TE
Tennis (TE)	4	4 MT, 6 NE, 2 SO	13	3 CR, 3 NE, 5 SO
Recall	83.9 %		68.3 %	
Precision	76.7 %		66.5 %	

that since each video class represents structured events unfolding in time, the appropriate class signatures are also present in the temporal domain such as shot-length, motion, TTC, and so forth. For example, though the plots of log shot-length of news, cricket and soccer were very similar, the frequent presence of wipe and high-motion distinguishes cricket and soccer from the news.

Sequence Classifier

As discussed before, the problem of sequence classification is to assign labels from one or more of the classes to the observation sequence, seq_T consisting of feature vectors $F^0, \ldots, F^T$. Table 2 shows the classification performance of the DBN models for six video classes. The experiment was conducted by training each DBN model with pre-classified sequences and testing with unclassified 30 sequences of the same class and 50 sequences of other classes. The DBN models for different classes had a different value of T, which was based on optimization performed during the training phase. The number of shots for DBN model of news was 5 shots, of soccer 8 shots, and of commercials 20 shots. The "news" class consists of all the segments, that is, newscaster shots, outdoor shots, and so forth. The commercial sequences all have a black frame at the end.

There are two paradigms of training the DBN: (i) Through a fixed number of shots (around seven to 10), and (ii) through a fixed number of frames (400 to 700 frames). The first paradigm works better than the second one as is clear from the recall and precision rates. This could be due to two reasons: first, having a fixed number of frames does not yield proper models for classes, and second, the DBN output is in terms of likelihood, which reduces as there are a larger number of shots in a sequence. Large numbers of shots implies more state transitions, and since each transition has probability of less than one, the overall likelihood decreases. Thus, classes with longer shotlengths are favored in the second paradigm, leading to misclassifications.

In general, DBN modeling gave good results for all the classes except soccer. It is interesting to note that commercials and news were detected with very high precision because of the presence of the characteristic black frame and the absence of high-motion, respectively. A large number of fade-in and fade-out effects are present in the MTV and commercial classes, and dissolves and wipes are frequently present in sports. The black frame feature also prevents the MTV class being classified as a commercial, though both of them have similar shot lengths and shot transitions.

The poor performance of the soccer class could be explained by the fact that the standard deviations of the DBN parameters corresponding to motion or shot length features were large, signifying that the degree of characterization of the soccer class was poor by features such as shot length. On the other hand, cricket was much more structured, especially with a large number of wipes.

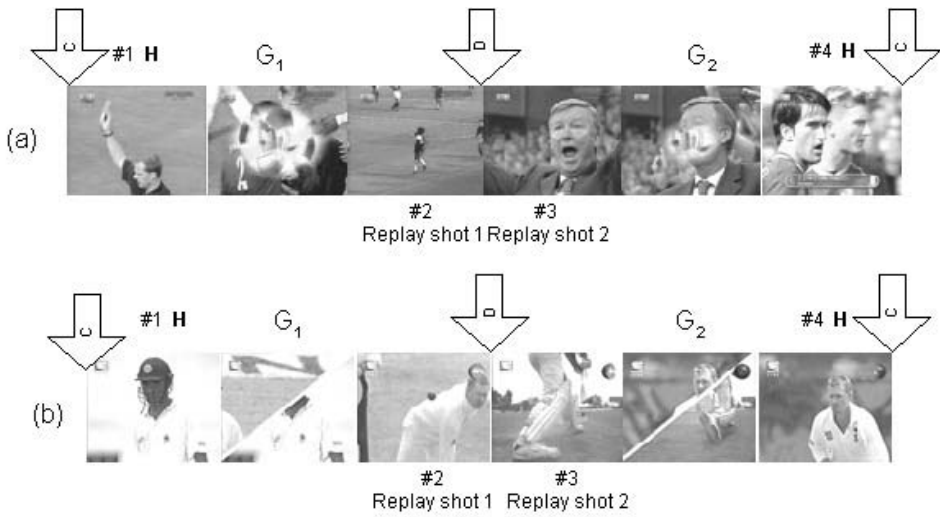
Highlight Extraction

In replays, the following conventions are typically used:

1. Replays are bounded by a pair of identical gradual transitions, which can either be a pair of wipes or a pair of dissolves.
2. Cuts and dissolves are the only transition types allowed between two successive shots during a replay. Figure 7 shows two interesting scenes from the soccer and cricket videos. Figure 7(a) shows a soccer match in which the player touched the football with his arm. A cut to a close-up of the referee showing a yellow card is used. A wipe is employed in the beginning of the replay, showing how the player touched the ball. A dissolve is used for showing the expression of the manager of the team, followed by a wipe to indicate the end of the replay. Finally, a close-up view of the player who received the yellow card is shown.

For the purpose of experiments, only two sports, cricket and soccer, were considered, although the same idea can be extended to most of the sports, if not all. This is because the interesting scene analysis is based on detecting the replays, the structure of which is the same for most sports.

Figure 7. Typical formats of the transition effects used in the replay sequences of (a) soccer match (b) cricket match



Below is a typical format of the transition effects around an interesting shot in cricket and soccer.

$\alpha E . \rightarrow C \rightarrow G1 \rightarrow . \beta\ E \rightarrow G2 \rightarrow E$
where,
E €{Cut, Dissolve}
C € {Cut}
G1, G2 € {Wipe, Dissolve}, G1 and G2 are the gradual transitions before and after the replay
D € {Dissolve}
α, β, γ *€{Natural number}*
$\rightarrow$ *is a ' followed by operator.*

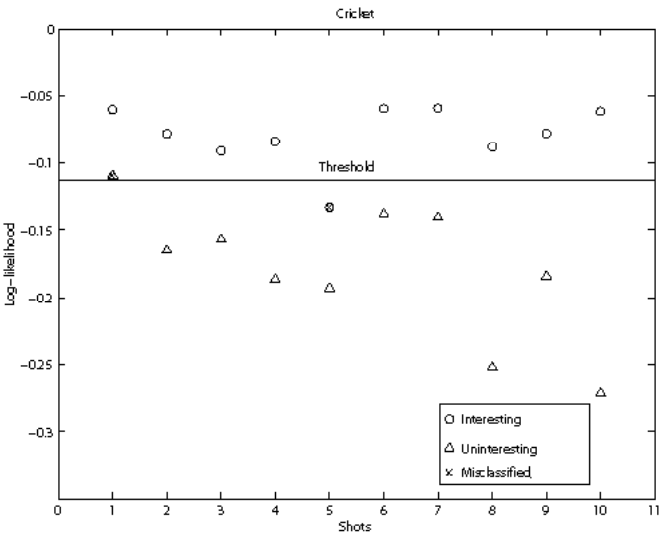
This special structure of replay is used for identifying interesting shots. The training sequence for both sports, which was typically 5 to 7 shots, consisted of interesting shots followed by the replays. Figure 8 shows the classification performance of DBN for highlight extraction. A threshold could be chosen on the likelihood returned by DBN based on training data (such that the threshold is less than the likelihood of most of the training sequences). All sequences possessing a likelihood more than the threshold of a DBN model are assigned an "interesting" class label. Figure 8 shows that only two misclassifications result with the testing. One recommendation could be to lower the threshold such that no interesting sequences are missed; although a few uninteresting sequences may also be labeled and retrieved to the user. Once an interesting scene is identified, the replays are removed before presenting it to the user.

An interesting detail in this application is that one might want to show the scoreboard during the extraction of highlights (scoreboards are generally preceded and followed by many dissolves). DBN can encode in its learning both the previous model of the shot followed by a replay and the scoreboard sequences with their characteristics.

Climax Characterization and Censoring Violence

Climax is the culmination of a gradual building up of tempo of events, such as chase scenes, which end up typically in violence or in passive events. Generally, a movie with a lot of climax scenes is

Figure 8. Classification performance for highlight extraction



classified as a "thriller" while a movie which has more violence scenes is classified as "action." Sequences from "thriller" or "action" movies could be used to train a DBN model, which can learn about the climax or violence structures easily. This application can be used to present a trailer to the user, or to classify the movie into different media categories and restrict the presentation of media to only acceptable scenes.

Over the last two decades, a large body of literature has linked the exposure to violent television with increased physical aggressiveness among children and violent criminal behavior (Kopel, 1995, p. 17; Centerwall, 1989). A modeling of the censoring process can be done to restrict access to media containing violence. Motion alone is insufficient to characterize violence, as many acceptable classes (especially sports like car racing, etc.) also possess high motion. On the other hand, shot-length and especially TTC are highly relevant due to the fact that the camera is generally at a short distance during violence (thus the movements of the actors yield impression of impending collisions). The cutting rate is also high, as many perspectives are generally covered. The set of the features used, though, remains the same as a sequence classifier application. Figure 10 shows a few images from a violent scene of a movie.

Figure 9 shows the classification performance for a censoring application. For each sequence in the training and test database, the opinions of three judges were sought in terms of one of the three categories: "violent", "nonviolent" and "cannot say." Majority voting was used to decide if a sequence is violent. The test samples were from sports, MTV, commercials and action movies. Figure 9 shows that while the violent scenes were correctly detected, two MTV shots were misclassified as "violent." For one sequence of these MTV sequences, the opinion of two judges was "cannot say," while the third one classified it as "violent" (Figure 11). The other sequence consisted of too many objects near to the camera with a large cutting rate although it should be classified as "nonviolent."

Figure 9. Classification performance for violence detection

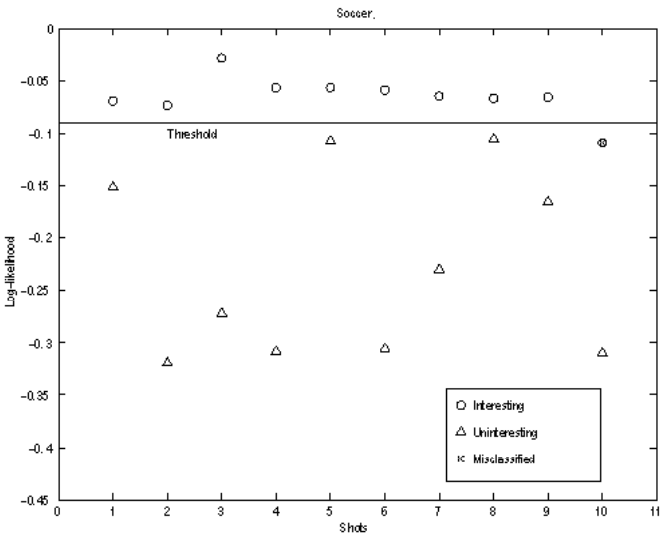


Figure 10. A violent scene that needs to be censored

Figure 11. An MTV scene which was misclassified as violent. There are many moving objects near the camera and the cutting rate is high.

DISCUSSION AND CONCLUSION

In this chapter, the integration of temporal context information and the modeling of this information through DBN framework were considered. Modeling through DBN removes the cumbersome task of manually designing a rule-based system. The design of such a rule-based system would have to be based on the low-level details, such as thresholding; besides, many temporal structures are difficult to observe but could be extracted by automatic learning approaches. DBN assignment of the initial labels on the data prepares it for the subsequent passes, where expert knowledge could be used without much difficulty.

The experiments conducted in this chapter employed a few perceptual-level features. The temporal properties of such features are more readily understood than low-level features. Though the inclusion of low-level features could enhance the characterization of the categories, it would raise the important issue of dealing with the high-dimensionality of the feature-space in the temporal domain. Thus, the extraction of information, which involves learning, decoding and inference, would require stronger and more efficient algorithms.

REFERENCES

Aigrain, P., & Joly, P. (1996). Medium knowledge-based macro-segmentation of video into sequences. *Intelligent Multimedia Information Retrieval.*

Boyen, X., Firedman, N., & Koller, D. (1999). Discovering the hidden structure of complex dynamic systems. *Proceedings of Uncertainty in Artificial Intelligence* (pp. 91-100).

Bryson, N., Dixon, R.N., Hunter, J.J., & Taylor, C. (1994). Contextual classification of cracks. *Image and vision computing, 12,* 149-154.

Centerwall, B. (1989). Exposure to television as a risk factor for violence. *Journal of Epidemology,* 643-652.

Chang, S. F., & Sundaram, H. (2000). Structural and semantic analysis of video. *IEEE International Conference on Multimedia and Expo* (pp. 687-690).

Dorai, C., & Venkatesh, S. (2001). Bridging the semantic gap in content management systems: Computational media aesthetics. *International Conference on Computational Semiotics in Games and New Media* (pp. 94-99).

Frith, U., & Robson, J. E. (1975). Perceiving the language of film in Perception, *4,* 97-103.

Garg, A., Pavlovic, V., & Rehg, J. M. (2000). Audio-visual speaker detection using Dynamic Bayesian networks. *IEEE Conference on Automatic Face and Gesture Recognition* (pp. 384-390).

Ghahramani, Z. (1997). Learning dynamic Bayesian networks. *Adaptive Processing of Temporal Information. Lecture Notes in AI.* SpringerVerlag.

Hamid, I. E., & Huang, Yan. (2003). Argmode activity recognition using graphical models. *IEEE CVPR Workshop on Event Mining: Detection and Recognition of Events in Video* (pp. 1-7).

Hummel, R. A., & Zucker, S. W. (1983). On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5,* 267-287.

Intille, S. S., & Bobick, A. F. (1995). Closed-world tracking. *IEEE International Conference on Computer Vision* (pp. 672-678).

Jain, A. K., Vailaya, A., &Wei, X. (1999). Query by video clip. *Multimedia Systems,* 369-384.

Jeon, B., & Landgrebe, D. A. (1990). Spatio-temporal contextual classification of remotely sensed multispectral data. *IEEE International Conference on Systems, Man and Cybernetics* (pp. 342-344).

Kettnaker, V. M. (2003). Time-dependent HMMs for visual intrusion detection. *IEEE CVPR Workshop on Event Mining: Detection and Recognition of Events in Video.*

Kittler, J., & Foglein, J. (1984). Contextual classification of multispectral pixel data. *Image and Vision Computing, 2,* 13-29.

Kopel, D. B. (1995). Massaging the medium: Analyzing and responding to media violence without harming the first. *Kansas Journal of Law and Public Policy, 4,* 17.

Kuleshov, L. (1974). *Kuleshov on film: Writing of Lev Kuleshov.* Berkeley, CA: University of California Press.

Meyer, F. G. (1994). Time-to-collision from first-order models of the motion fields. *IEEE Transactions of Robotics and Automation* (pp. 792-798).

Mittal, A., & Altman, E. (2003). Contextual information extraction for video data. *The 9th International Conference on Multimedia Modeling (MMM),* Taiwan (pp. 209-223).

Mittal, A., & Cheong, L.-F. (2001). Dynamic Bayesian framework for extracting temporal structure in video. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2,* 110-115.

Mittal, A., & Cheong, L.-F. (2003). Framework for synthesizing semantic-level indices. *Journal of Multimedia Tools and Application,* 135-158.

Nack, F., & Parkes, A. (1997).The application of video semantics and theme representation in automated video editing. *Multimedia tools and applications,* 57-83.

Nicholson, A. (1994). Dynamic belief networks for discrete monitoring. *IEEE Transactions on Systems, Man, and Cybernetics, 24*(11), 1593-1610.

Olson, I. R., & Chun, M. M. (2001). Temporal contextual cueing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Pavlovic, V., Frey, B., & Huang, T. (1999). Time-series classification using mixed-state dynamic Bayesian networks. *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 609-615).

Pavlovic, V., Garg, A., Rehg, J., & Huang, T. (2000). Multimodal speaker detection using error feedback dynamic Bayesian networks. *IEEE Conference on Computer Vision and Pattern Recognition.*

Pynadath, D. V., & Wellman, M. P. (1995). Accounting for context in plan recognition with application to traffic monitoring. *International Conference on Artificial Intelligence, 11.*

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected application in speech recognition. *Proceedings of the IEEE* (vol. 77, pp. 257-286).

Sondhauss, U., & Weihs, C.(1999). Dynamic bayesian networks for classification of business cycles. SFB Technical report No. 17. Online at *http://www.statistik.uni-dortmund.de/*

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimal

decoding algorithm. *IEEE Transactions on Information Theory* (pp. 260-269).

Yeo, B. L., & Liu, B.(1995). Rapid scene analysis on compressed video. *IEEE Transactions on Circuits, Systems, and Video Technology* (pp. 533-544).

Yu, T. S., & Fu, K. S. (1983). Recursive contextual classification using a spatial stochastic model. *Pattern Recognition, 16,* 89-108.

Zweig, G. G. (1998). *Speech recognition with dynamic Bayesian networks.* PhD thesis, Dept. of Computer Science, University of California, Berkeley.

This work was previously published in Managing Multimedia Semantics, edited by U. Srinivasan and S. Nepal, pp. 77-98, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 2.15
Design Principles for Active Audio and Video Fingerprinting

Martin Steinebach
Fraunhofer IPSI, Germany

Jana Dittmann
Otto-von-Guericke-University Magdeburg, Germany

ABSTRACT

Active fingerprinting combines digital media watermarking and codes for collusion-secure customer identification. This requires specialized strategies for watermark embedding to lessen the thread of attacks like marked media comparison or mixing. We introduce basic technologies for fingerprinting and digital watermarking and possible attacks against active fingerprinting. Based on this, we provide test results, discuss the consequences and suggest an optimized embedding method for audio fingerprinting.

INTRODUCTION

Robust digital watermarking is the enabling technology for a number of approaches related to copyright protection mechanisms: Proof of ownership on copyrighted material, detection of the originator of illegally made copies and monitoring the usage of the copyrighted multimedia data are typical examples where watermarking is applied. A general overview about digital watermarking can be found in a variety of existing publications, for example in Cox, Miller and Bloom (2002) or Dittmann, Wohlmacher and Nahrstedt (2001).

While stopping the reproduction of illegal copies may be the first goal for copyright holders, discouraging pirates from distributing copies is the more realistic goal today. It can be observed that current copy protection or digital rights management systems tend to fail in stopping pirates (Pfitzmann, Federrath, & Kuhn, 2002). One important reason for this is the fact that media data usually leave a controlled digital environment when they are consumed, enabling analogue copies of high quality of material protect with digital mechanisms. Under these circumstances, the challenge is to find the most discouraging method making it especially dangerous for pirates

to distribute copies. Identification of a copyrighted work by embedding a watermark or retrieving a passive fingerprint (Allamanche, Herre, Helmuth, Fröba, Kasten, & Cremer, 2001) is necessary for preventing large-scale production of illegal CD or DVD copies, but does not stop people from distributing single copies or uploading them to file-sharing networks. Here a method that enables the copyright holder to trace an illegal copy to its source would be much more effective, as pirates would loose their anonymity and therefore have to fear detection and punishment.

Embedding unique customer identification as a watermark into data to identify illegal copies of documents is called fingerprinting. Basically, watermarks, labels or codes embedded into multimedia data to enforce copyright must uniquely identify the data as property of the copyright holder. They also must be difficult to be removed, even after various media transformation processes. Thus the goal of a label is to always remain present in the data. Digital fingerprinting, which embeds customer information into the data to enable detection of license infringement, raises the additional problem that we produce different copies for each customer. Attackers can compare several fingerprinted copies to find and destroy the embedded identification string by altering the data in those places where a difference was detected.

In this chapter, we introduce a method for embedding customer identification into multimedia data: Active digital fingerprinting is a combination of robust digital watermarking and the creation of a collision-secure customer vector. In literature we also find the term *collusion secure fingerprinting* or *coalition attack secure fingerprinting.* There is also another mechanism often called fingerprinting in multimedia security, the identification of content with robust hash algorithms; see for example in Haitsma, Kalker and Oostveen (2001). To be able to distinguish both methods, robust hashes are called *passive fingerprinting* and collision-free customer identification watermarks are called *active fingerprinting.* Whenever we write fingerprinting in this chapter, we mean active fingerprinting.

MOTIVATION

To achieve customer identification directly connected to the copy of the media to be protected, embedding a robust watermark with a customer identification number—called ID from hereon—is a first solution. This prevents the removal of the ID in most cases, as watermarking algorithms become robust to most media processing, like lossy compression or DA/AD conversion.

A simple example:

The content provider wants to sell four copies of an audio file to his customers. To be able to trace the source of an illegal distribution of the file, he embeds a different bit sequence in each copy. For customer A, he embeds "00," for B "01," for C "10" and for D "11".

If he finds a copy of the audio file only sold to those four customers, he could try to retrieve the watermark from the copy. As he uses a robust watermarking algorithm, he is able to find the watermark and to identify the source. For example, he detects the watermark "01" and concludes the source is B. In the case of very strong attacks—which would reduce the quality and make the copies less attractive—he may not be able to detect the watermark, but when the copy is of little value, he does not worry about this.

But due to watermarking characteristics, a much more dangerous situation can occur: Imagine A and D know each other and want to distribute illegal copies of the audio file. They know the file is watermarked and have a certain level of understanding regarding this technology. Therefore they compare both copies to each other, showing differences at certain positions. Knowing most watermarking algorithms can be confused

in this way, they now mix both copies, creating a copy consisting of both customers' copies.

This could render the watermarking algorithm unable to detect the watermarking information embedded. The illegal copy would be of good quality but still not traceable. Even worse, this could lead to pointing at a third customer who is innocent: If A's "00" and D's "11" are mixed, depending on the attacking algorithm and the watermarking method, it can happen that "01" or "10" is detected and B or C are accused.

A possible solution to this problem is to embed checksums together with the customer ID, significantly reducing the probability of false accusations, as the randomly generated new watermark will not fit to the checksum and the watermark becomes useless. Still, no customer identification takes place and the attack is successful. Therefore, a collusion-robust method for customer identification is required.

BASIC TECHNOLOGIES

Before we discuss optimisation methods for active digital fingerprinting, we need to provide an overview to customer identification codes and to identify requirements regarding digital watermarking in this scenario.

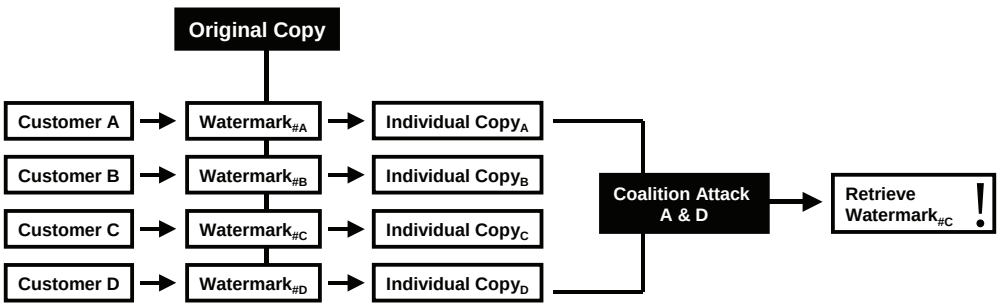
From Dittmann, Behr, Stabenau, Schmitt, Schwenk and Ueberberg (1999), a digital fingerprinting scheme consists of:

- a number of marking positions in the document
- a watermarking embedder to embed letters from a certain alphabet—most often bits—at these marking positions
- a fingerprint generator, which selects the letters to be embedded for each marking position depending on the customer
- a watermarking detector to retrieve a watermark from a marked copy
- a fingerprint interpreter, which outputs at least one customer from the retrieved watermarking information

Different copies of a document containing digital fingerprints differ at most at the marking positions. An attack — as already described — to remove a fingerprint therefore consists of comparing two or more fingerprinted documents and altering these documents randomly in those places where a difference was detected. If three or more documents are compared, a majority decision can be applied to improve this kind of attack: For the area where the documents differ, one will choose the value that is present in most of the documents. The only marking positions the pirates cannot detect are those positions that contain the same letter in all the compared documents. We call the set of these marking positions the intersection of the different fingerprints.

The major challenge of active digital fingerprinting is to create a sequence of bits (or letters)

Figure 1. Coalition attack scheme



that is robust against these comparisons. Even if the attackers can identify differences between their copies, mixing the copies must not lead to a copy in which none of the attackers can be identified.

Active Fingerprinting

To solve the problem of the coalition attack, we use the Boneh-Shaw fingerprint and the Schwenk-Ueberberg fingerprint algorithm (Boneh & Shaw, 1995; Dittmann et al., 1999). Both algorithms offer the possibility to find the customers who have committed the coalition attack. As an application example, we have applied both schemes in a video fingerprinting solution with coalition resistance in Dittmann, Hauer, Vielhauer, Schwenk and Saar (2001) and an analysis of the resistants of audio watermarking in Steinebach, Dittmann and Saar (2002). In the following two subchapters we summarize the two fingerprinting schemes.

Schwenk Fingerprint Scheme

The Schwenk et al. approach (Dittmann et al., 1999) puts the information to trace the pirates into the intersection of up to *d* fingerprints. This allows us in the best case (e.g., automated attacks like computing the average of fingerprinted images) to detect all pirates. In the worst case (removal of individually selected marks), we can detect the pirates with a negligibly small one-sided error probability; that is, we will never accuse innocent customers.

The fingerprint vector is spread over the marking positions. The marking positions for each customer are the same in every customer copy and the intersection of different fingerprints can therefore not be detected. With the remaining marked points, the intersection of all used copies, it is possible to follow up on all customers who have worked together. Another important parameter is the number *n* of copies that can be generated with such a scheme. The scheme uses techniques from finite projective geometry (Beutelspacher & Rosenbaum, 1998; Hirschfeld, 1998) to construct d-detecting fingerprinting schemes with q+1 possible copies. This scheme needs $n=q^d+q^{d-1}+...+q+1$ marking positions in the document. As we see, this can be a huge length and can cause problems with the capacity of the watermarking scheme. The idea to build the customer vector is based on finite geometries and the detailed mathematical background will be provided in the final section.

Boneh-Shaw Fingerprint Scheme

The scheme of Boneh and Shaw (1995) is also used to recognize the coalition attack, but it is another scheme. Here it is noticeable that we do not necessarily find all pirates, with a (any arbitrary small) probability *e* that we get the wrong customer, and each fingerprint has a different number of zeros.

The number of customers is *q* and with *q* and *e* you can get the repeats *d*. The fingerprint vector consists of (*q*-1) blocks of the length d ("d-blocks"), and the total length of the embedded fingerprint computes as *d**(*q*-1). Depending on the repeats the customer vector can be very long and cause problems with the capacity of the watermarking algorithm. The idea to build the fingerprinting vector for each customer is simple: The first customer has the value one in all marked points; for the second customer all marked points without the fist "d-block" are ones; in the third all marked points without the first two "d-blocks" are ones, and so forth. The last customer has the value 0 in all marked points.

With a permutation of the fingerprint vector we get a higher security, because the pirates can find differences between the copies, but they cannot assign it to a special d-block. In the final version we provide the detailed mathematical background.

Digital Watermarking

Digital watermarking is in general a method of embedding information into a cover file. In our case, the cover consists of audio or video files. Depending on the application scenario, the information embedded will differ. Even the basic concept of the watermarking algorithms may change, as there are robust, fragile and invertible watermarking schemes (Cox et al., 2001; Dittmann, 2000; Dittmann et al., 2001; Petticolas & Katzenbeisser, 2000). For copyright protection, usually robust watermarking is applied. Still, numerous requirements need to be identified to adjust a watermarking algorithm to a specific scenario. In this section, we discuss the watermarking requirements with respect to the active fingerprinting application.

Fingerprinting can only take place if the customer is known. This is the case in, for example, Web shop environments, more generally speaking in on-demand-scenarios that require customer authentication. Copies of songs ripped from CDs bought anonymously in stores cannot be marked this way. Embedding a fingerprint in an on-demand situation induces a number of requirements to the watermarking algorithms:

- *Transparency* is a common requirement for marking digital media in e-commerce environments, as the quality of the content acting as a cover for the watermark must not be reduced.
- *Robustness* is necessary against common media operations like lossy compression and format changes.
- *Payload* must be high enough to include the fingerprint, which usually consists of a long bit vector. This can become a critical requirement.
- *Security* is of special importance in this case, as the existence of several copies of the same cover with different embedded fingerprints enables a number of specialized attacks commonly called *coalition attacks.*
- *Complexity* has to be low enough to enable online and real-time marking. A customer who wants to download a song is not willing to stay online for a long time until his or her personalized copy is available. As there will be multiple customers at the same time and media data may have a playing time of an hour or more, either streaming concepts or multiple real-time embedding speed will be necessary. Furthermore, in most cases non-blind methods are suitable where the original is needed during retrieval or detection.
- *Verification* should be performed in a secret environment. The content provider uses a secret watermarking key to embed and retrieve the watermarks. Customers do not know this key as attackers could easily verify their success with it.

This scenario-specific list of parameters shows the difference to common copyright protection environments: While robustness and transparency are of similar interest, in our scenario payload, security and complexity become more important. As active fingerprinting will only be applied if the content provider has to be prepared for attacks against more simple customer identification schemes, security is of special interest. One can assume that specialized attacks against the watermarks will take place. Complexity needs to be low in comparison to embedding a copyright notice, as in active fingerprinting each copy sold needs to be watermarked. This easily can become a bottleneck if the algorithms are not designed accordingly. In general, active fingerprints are much longer than copyright notes, inducing higher payload requirements in our scenario.

To summarize these observations, it becomes clear that not all watermarking algorithms suitable for robust copyright watermarking will be equally suited for active fingerprinting. Only those algorithms that provide high security, high

payload and a low complexity in addition to a high transparency and good robustness may be chosen as a watermarking method.

ADJUSTING WATERMARKING ALGORITHMS TO ACTIVE FINGERPRINTING

To apply active fingerprinting in tracing illegal copies we need a digital watermarking algorithm. Current digital watermarking techniques may embed the generated fingerprinting information redundantly and randomly over the media file. With a random distribution, the intersection of the proposed fingerprints may be destroyed by coalition attacks. Therefore it is important to ensure that one bit with a certain position in the fingerprint vector is always embedded at the same place in the media file for every copy. Only then an intersection is undetectable for an attacker.

Audio Watermarking

To use the properties of the fingerprinting mechanisms to identify the customers who attacked the watermark we build a watermarking scheme with a fixed number of marking positions in each copy of the audio file (Steinebach et al., 2002). These marking positions can be selected based on a secret key and a psycho-acoustic model to find secure and transparent positions. The fingerprinting algorithm generates the fingerprint vector over the binary alphabet {0,1}. The watermarking algorithm embeds this binary vector at the chosen marking positions.

Watermarking algorithms use different methods to embed a message into a cover. The way the message is embedded is relevant for the security of the watermarking and fingerprinting combination: A PCM audio stream consists of a sequence of audio samples over time. Our algorithm uses a group of successive samples, for example, 2048, to embed a single bit of the complete message. Figure 2 illustrates this: The bit sequence 01011 is embedded in a 1-second audio segment by separating the audio into groups of samples and embedding one bit in each of the segments.

This leads to the following situation: If two different bit vectors are embedded in two copies of the same cover with the same key, the two copies differ exactly in those segments where different bits have been embedded as information. Figure 3 shows two embedded bit vectors “01011” and

Figure 2. Audio watermarking over time

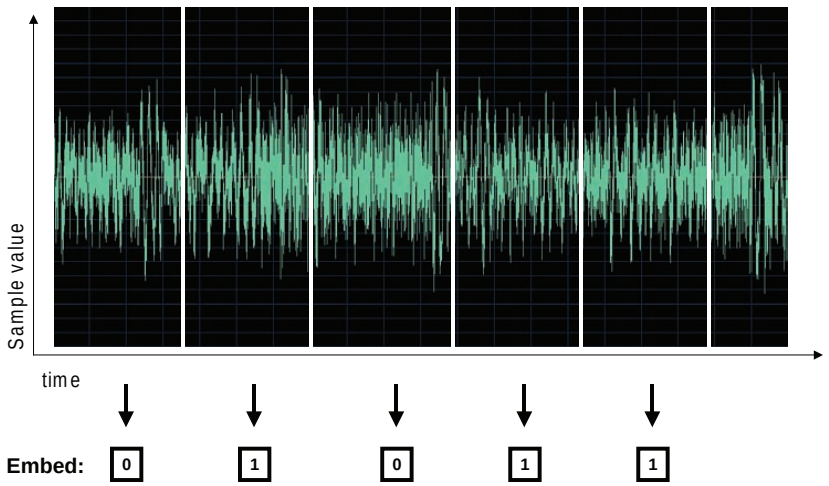
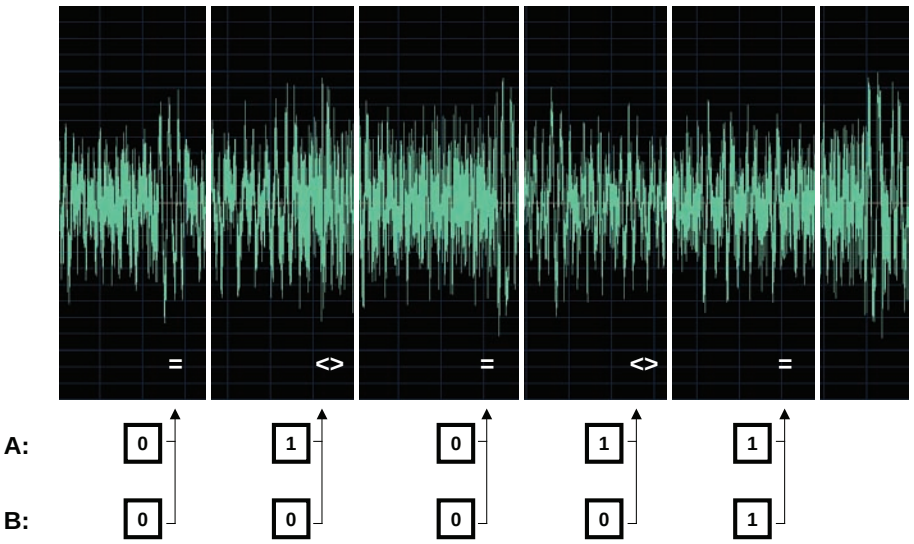


Figure 3. Different embedded bit vectors lead to different segments in the copies



"00001". Both have been embedded in the same cover audio file. If A and B compare their copies, they find equal segments at positions 1, 3 and 5 and different segments at positions 2 and 4.

Video Watermarking

In Dittmann et al. (2001) we have introduced for Schwenk Fingerprint Scheme and the Boneh-Shaw Fingerprint Scheme, a video fingerprinting solution and the coalition resistance. To mark the video, we generate positions within the frame to embed the watermark information (in the video the positions stand for scenes). Each customer has his or her own fingerprint, which contains a number of "1" and "0". Each fingerprint vector is assigned to marking positions in the document to prevent the coalition attack. The only marking positions the pirates cannot detect are those positions that contain the same letter in all the compared documents. We call the set of these marking positions the intersection of the different fingerprints.

Three general problems emerge during the development of the watermark (Dittmann et al., 2001):

- *Robustness.* To improve the robustness against the coalition attack, we embed one fingerprint vector bit in a whole scene. So we reach a resistance against statistical attacks, like average calculation of look alike frames. With this method we can make the frame cutting and frame changing ineffective. We have not contemplated the cutting of a whole scene yet. In the current prototype we mark a group of pictures GOP for one fingerprint bit. We add a pseudo-random sequence to the first AC values of the luminance DCT blocks of an intracoded macroblock in all I-Frames of the video.
- *Capacity.* The basis of the video watermark is an algorithm, which was developed for still images (Dittmann et al., 1999). In still images the whole fingerprint is embedded into the image and the capacity is restricted. With the I-Frame in a video, the capacity is much better. To achieve high robustness, we embed one watermark information bit into a scene. Thus the video must have a minimal length. Additional to the embedding of the watermark, the data rate can increase. The

problem of synchronization between the audio and video stream can arise or data rate is raised. Basically, with the embedding of the watermark we must synchronize the audio and video stream.

- *Transparence*. To improve the transparency we use a visual model. With the visual model the watermark strength is calculated for every marking position individually. Additionally, we use the same marking position for each frame.

DESIGNING ACTIVE FINGERPRINTING ALGORITHMS

Combining customer fingerprints and existing robust watermarking algorithms to provide active fingerprinting is only a first approach to solve the challenge of customer tracking. While existing algorithms may offer parameters to optimise them for this application, new algorithms especially designed for this purpose may lead to superior performance in this domain. In this section, we discuss approaches on digital watermarking algorithm design for active fingerprinting.

Fingerprinting-Optimised Audio Watermarking

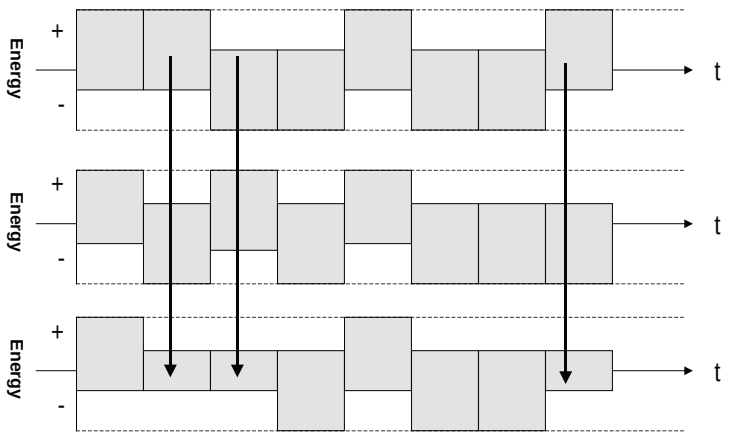
For identifying users that took part in a coalition attack, it could be helpful to change the embedding algorithm so that a rule could be set for mixing two fingerprints. If every time an embedded "0" and "1" are mixed, one specific bit occurs, we would receive a bit vector much more easy to interpret. In the case of the Schwenk algorithm, mixing a "0" and a "1" should always result in a "0" as the "1"s are used to identify the group of attackers.

An example:

The fingerprints A and B differ at position 2 and 5. This leads to 2^2 possible results of a fingerprint attack. Figure 4 shows the reason for this behaviour: Both 0 and 1 are embedded at equal strength. A coalition attack results in traces of both bits at similar strengths at the same position. The watermark detector will therefore have a comparatively random bit at these positions.

After an optimisation for the Schwenk fingerprint the only possible result of a coalition attack with the fingerprints A and B from the example above should be "0010001" identifying both attackers by the shared "1"s at position 3 and 6.

Figure 4. Even-strength watermarking leads to undeterminable results after coalition attacks



At the positions 2 and 5 where the bit values of both fingerprints differ, both times the "0" was dominant in the attack.

This characteristic can be achieved by using different embedding strengths for both bits. In the case of middle or mix attacks this would result in the bit embedded with more strength surviving the coalition attacks. Figure 5 illustrates this concept. Bit values are embedded as a positive or negative energy. Now if we embed a bit, we use more energy for one bit type then the other. When the two energy levels are later mixed by a coalition attack, the energy type embedded with more strength is dominant. For the Schwenk algorithm, the bit 0 would be embedded with more energy then bit 1.

In Figure 5, the positive embedding energy is stronger then the negative one. In the last row the result of a coalition attack is shown: Whenever a positive and a negative energy position is mixed, the result is positive and the retrieved watermarking bit can be predetermined.

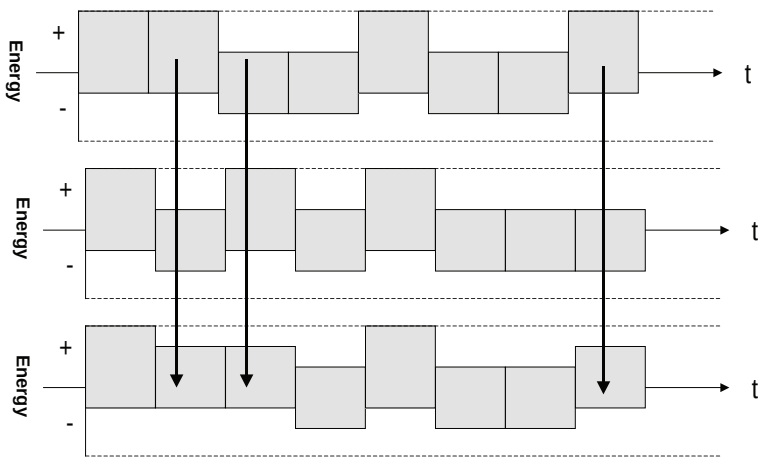
EXAMPLE SCENARIO: CINEMA APPLICATION

Fingerprinting media files today is seen as a promising way of discouraging illegal transfers of copyrighted material. Therefore example scenarios for this technology come from media distribution, especially where a small number of copies exist but leaking of these copies to the public results in major damage.

One appropriate example is the distribution of movies: In recent times copies of movies often happen to be available via Internet as illegal copies before or at the same day they are shown in cinemas. This leads to two possible leaks in distribution:

1. If the movie is available before it is shown in the cinema, some promotional copy of the movie may have been used as a master.
2. If the movie is available right when shown in the cinema, someone may have recorded it with a small video camera.

Figure 5. Different watermarking strengths for 0 and 1 lead to predetermined results after coalition attacks



Tracing illegal copies is more difficult in (2) then in (1). The two leaks strongly differ with regards to the watermarking parameters:

- Robustness in (1) is only necessary against digital video format change if the promotion copy is on DVD, or against high-quality digitisation if the copy is on videotape. Leak (2) requires robustness against a low quality analogue to digital conversion, as the movie is recorded by a small digital camera in a noisy surrounding.
- Transparency, on the other hand, needs to be higher or at least more reliable in (2) than in (1), as a low audio or video quality caused by embedding the watermark will not be accepted by movie theatres and customers.
- The required payload of the watermark may also be higher in (2) than in (1), as one can assume there will be fewer promotional copies than actual movie copies for the cinemas. This leads to more individual customers to be identified by the fingerprints, making them significantly longer.

While, therefore leak (2) may be more challenging then (1), both can be addressed with the same strategy:

- Create a movie master
- Create the required amount of fingerprinted copies from this master
- Distribute the fingerprinted copies
- Search for occurring illegal copies
- Retrieve the fingerprint from the copy
- Identify the leak with the help of the fingerprint

Attacks Against Fingerprinted Copies

As a movie consists of video as well as audio information, watermarking algorithms for both media types can be used for fingerprint embedding. While the watermarking algorithms may be able to satisfy all the scenario-dependent requirements stated above, the fingerprint may also be subject to specialized attacks as soon as it becomes known to the public that fingerprinting is used for tracing copies. This is unavoidable if discouragement is desired.

Let us assume we fingerprint promotional DVDs for tracing leaks (1) using an MPEG video watermark. Two recipients of the promotional copies wanting to distribute illegal copies and willing to work together now can start coalition attacks to remove or corrupt the fingerprints.

The coalition security implies, for example, the following attacks:

(a) Attacks to separate frame areas
(b) Attacks to whole frame
(c) Attacks to whole scenes

The time and practical effort grows from (a) to (c). The video must be split in the important areas. Additionally there must be knowledge about the MPEG video format. The attack over whole frames, like the exchange of frames, is only possible with visually similar frames, because with different frames the semantics of the frames can be destroyed. Only for attacks over whole scenes the watermark has no robustness, because one bit of the fingerprint vector will be cut out. But with the cut off of whole scenes the semantics of the video will be decreased.

Optimisation Potential

To be less vulnerable against coalition attacks, we introduced a strategy for fingerprinting-optimised audio watermarking in this chapter. This strategy can be applied in the cinema-application if a certain loss of audio quality is acceptable, which may be the case in promotional copy distribution.

First test results with embedding the bits 0 and 1 of the audio watermark with different energy are promising. Depending on the energy difference, error rates after coalition attacks are reduced by up to 50%. Error rates have been calculated by counting

the number of times bit 1 has been replaced by bit 0 after a coalition attack.

Figure 6 shows that the reduction of error rates is related to the increase of energy difference between bit 0 and bit 1. If both are embedded at equal strength (0 dB difference), the error rate is above 50%. On the other hand, at a difference of 12 dB, almost no errors occurred.

If the quality loss caused by the strong watermark for bit 1 can be accepted, embedding watermark bits with differing energy seems to be an improvement regarding robustness against coalition attacks. In our example, to reduce the error rate below 10%, we would need an embedding difference of 6 dB, which produces a quality loss similar to mp3 encoding at 192 kbps. This should be acceptable for a huge number of applications.

SUMMARY AND CONCLUSIONS

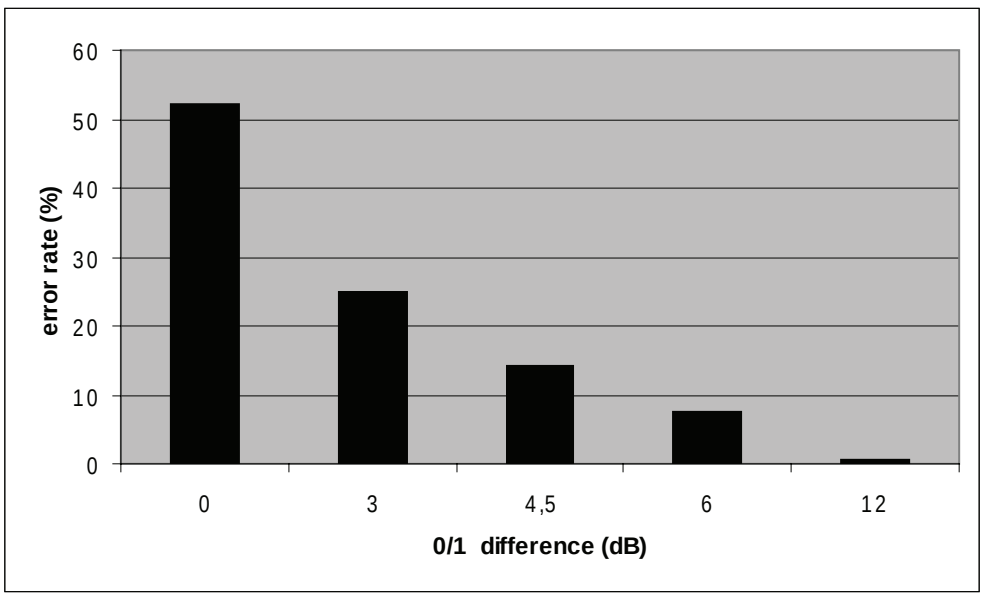
Altogether, digital watermarking to embed fingerprinting information is a pragmatic approach to discourage the illegal use of the copied data. The wide variety of existing watermarking algorithms reflects the business relevance. Beside the robustness to common media transformations, coalition attacks raise importance and become a critical factor for example in the design of cinema applications. As, for example, a special session "Cinema application" at SPIE 2003 shows, the combination of secure fingerprint schemes and the watermarking algorithms itself seems to be still an open problem. Reasons are in most cases the limited capacity for embedding the collusion secure fingerprint as well as synchronisation problems.

Besides the development of watermarking algorithms and collusion secure fingerprint vector design our future goal is to design interactive tools that strengthen the producers' acceptance to use digital watermarking techniques to offer their data in a more secure way in the digital marketplace.

REFERENCES

Allamanche, E., Herre, J., Helmuth, O., Fröba, B., Kasten, T., & Cremer, M. (2001). Content-based identification of audio material using MPEG-7 low

Figure 6. Error rates of fingerprints



level description. *Proceedings of the International Symposium of Music Information Retrieval.*

Beutelspacher, A., & Rosenbaum, U. (1998). *Projective geometry.* Cambridge University Press.

Boneh, D., & Shaw, J. (1995). Collusion-secure fingerprinting for digital data. *Proceedings of CRYPTO'95,* LNCS 963, (pp. 452-465). Springer.

Cox, I., Miller, M., & Bloom, J. (2002). *Digital watermarking,* ISBN 1-55860-714-5. San Diego, CA: Academic Press.

Dittmann, J. (2000). *Digitale wasserzeichen,* ISBN 3-540- 66661-3. Springer Verlag.

Dittmann, J., Behr, A., Stabenau, M., Schmitt, P., Schwenk, J., & Ueberberg, J. (1999). Combining digital watermarks and collusion secure fingerprints for digital images. *Proceedings of SPIE, 3657,* (pp. 3657-51). San Jose, CA: Electronic Imaging.

Dittmann, J., Hauer, E., Vielhauer, C., Schwenk, J., & Saar, E. (2001). Customer identification for MPEG video based on digital fingerprints. *Proceedings of Advances in Multimedia Information Processing - PCM 2001, The Second IEEE Pacific Rim Conference on Multimedia*, Beijing, China, ISBN 3-540-42680-9, (pp. 383-390). Berlin: Springer Verlag.

Dittmann, J., Wohlmacher, P., & Nahrstedt, K. (2001, October-December). Multimedia and security – Using cryptographic and watermarking algorithms. ISSN 1070-986X *IEEE MultiMedia, 8*(4), 54-65.

Haitsma, J., Kalker, T., & Oostveen, J. (2001). Robust audio hashing for content identification. *Proceedings of the Content-Based Multimedia Indexing.*

Hirschfeld, J.W.P. (1998). *Projective geometries over finite fields* (2nd ed.). Oxford University Press.

Petticolas, F., & Katzenbeisser, S. (2000). *Information hiding techniques for steganography and digital watermarking.* Artech House Computer Security Series, ISBN: 1580530354.

Pfitzmann, A., Federrath, J., & Kuhn, M. (2002). *DRM-studie dmmv- technischer teil.*

Steinebach, M., Dittmann, J., & Saar, E. (2002, September 26 - 27). Combined fingerprinting attacks against digital audio watermarking: Methods, results and solutions. In B. Jerman-Blazic & T. Klobucar (Eds.), *Proceedings of Advanced Communications and Multimedia Security, IFIP TC6/TC11 6th Joint Working Conference on Communications and Multimedia Security*, Portoroz, Slovenia (pp. 197 - 212, ISBN 1-4020-7206-6). Kluwer Academic Publishers.

This work was previously published in Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property, edited by C.-S. Lu, pp. 157-172, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.16
A Service-Oriented Multimedia Componentization Model

Jia Zhang
Northern Illinois University, USA

Liang-Jie Zhang
IBM T.J. Watson Research Center, USA

Francis Quek
Virginia Tech, USA

Jen-Yao Chung
IBM T.J. Watson Research Center, USA

ABSTRACT

As Web services become more and more popular, how to manage multimedia Web services that can be composed as value-added service solutions remains challenging. This paper presents a service-oriented multimedia componentization model to support Quality of Service (QoS)-centered, device-independent multimedia Web services, which seamlessly incorporates cutting-edge technologies relating to Web services. A multimedia Web service is divided into control flow and data flow. Each can be delivered via different infrastructures and channels. Enhancements are proposed to facilitate Simple Object Access Protocol (SOAP) and Composite Capability/Preference Profiles (CC/PP) protocols to improve their flexibility to serve multimedia Web services. We present a set of experiments that show the viability of our service-oriented componentization model that can support efficient delivery and management of multimedia Web services.

INTRODUCTION

Simply put, a Web service is a programmable Web application that is universally accessible through standard Internet protocols (Ferris, 2003). The rapidly emerging technology of Web services

exhibits the capability of facilitating business-to-business (B2B) collaboration in an unprecedented way. By means of each organization exposing its software services on the Internet and making them universally accessible via standard programmatic interfaces, this Web services paradigm enables and facilitates the sharing of heterogeneous data and software resources among collaborating organizations (Benatallah, 2002). In addition, Web services technology provides a uniform framework to increase cross-language and cross-platform interoperability for distributed computing and resource sharing over the Internet. Furthermore, this paradigm of Web services opens a new cost-effective way of engineering software to quickly aggregate individually published Web services as components into new services. Therefore, the Web services technology has attained significant momentum in both academia and industry.

If the sharable data to be published by a Web service contain multimedia content, which refers to information that seamlessly integrates multiple media types in a synchronized and interactive presentation, the Web service is considered as a multimedia Web service. Multimedia Web services pose new challenges due to the unique characteristics of multimedia data (Khan, 2002). First, the transport of the multimedia information has to meet some Quality of Service (QoS) requirements, such as the synchronization within and among different multimedia data streams or real-time delivery. For example, let us consider a typical Video on Demand (VoD) service, an Internet Kara OK service. It is critical to provide a significant short-response-time service to a VIP customer. In addition, the audio and video information needs to be synchronized on customer's system. Second, the Simple Object Access Protocol (SOAP), the core transport technique of Web services, does not support massive message transport that is imperative for multimedia content transport, or multimedia QoS requirements (Khan, 2002). Third, with the advancement of wireless information appliances, Web service interfaces provide a means to enable the content or service to be created once and accessed by multiple SOAP-enabled [4-6] devices, such as wireless phones (NORTEL), Personal Digital Assistance (PDAs), set-top boxes, as well as regular Web browsers. A Web service is thus considered to be device independent if it can be delivered to different devices (Han, 2000). How to deliver a multimedia Web service to users based upon their possessed devices remains challenging.

In summary, the interoperability of multimedia Web services is not without penalty since the value added by this new Web service paradigm can be largely defeated if a multimedia Web service: (1) cannot guarantee QoS attributes; (2) cannot be transported via the Internet in an organized manner; and (3) cannot be effectively adapted to end devices including mobile devices. In this paper, we present a solution to these existing issues. We accomplish this goal in several ways. First, we propose a separation of control flow and data flow for multimedia Web services, using SOAP to transport the control flow. Second, we propose enhancements to SOAP to serve the transportation of multimedia Web services. Third, we propose enhancements to Composite Capability/Preference Profiles (CC/PP) protocol (CCPP) to provide an easy and flexible way to split and adapt multimedia Web services to appropriate composite devices, and increase the flexibility for users to manage multi-devices. Finally, we propose a service-oriented multimedia componentization model to support device-independent multimedia Web services.

This paper is organized as follows. We first briefly introduce some core techniques of multimedia Web services and related work. Then, we present our solution and we present performance analysis. Finally, we summarize the contributions and innovations, assess limitations and discuss future work directions.

CORE TECHNIQUES OF MULTIMEDIA WEB SERVICES

In this section, we will first briefly introduce the core techniques and standards of multimedia Web services to provide readers with some background context.

Web services typically adopt a provider/broker/requester architectural model (Roy, 2001). A service provider registers Web services at service brokers; a service broker publishes registered services; and a service requester searches Web services from service brokers. The essential aspect of this model is the concept of dynamic invocation: Web services are hosted by service providers, and service requesters dynamically invoke the Web services over the Internet on an on-demand basis.

In order to enable the communications among service providers, service brokers, and service requesters, the paradigm of Web services mainly embraces three core categories of supporting facilities: communication protocols, service descriptions, and service discovery (Roy, 2001). Each category possesses its own *ad hoc* standard. The Simple Object Access Protocol (SOAP) acts as a simple and lightweight protocol for exchanging structured and typed information among Web services. The Web Service Description Language (WSDL) is an XML-based description language that is used to describe the programmatic interfaces of Web services. The Universal Description, Discovery, and Integration (UDDI) standard provides a mechanism to publish, register, and locate Web services. It should be noted that here we adopt a narrow definition of Web services that refers to an implicit definition of SOAP+WSDL+UDDI for the purpose of simplicity. This implies a focus on the management of stand-alone Web services, instead of the compositions of and the interactions among multiple Web services.

As more business organizations adopt Web services technology to publish their sharable data to make their services accessible to more other organizations, it is possible that the data to be published include multimedia information, for example, audio and video. Because multimedia data exhibits unique features that require specific handling (Khan, 2002), it is necessary to examine the existing Web services techniques to better support multimedia Web services. When the SOAP+WSDL+UDDI framework is applied to support a multimedia Web service, two major factors influence the success of a multimedia Web service: (1) the transport of multimedia information over the Web, and (2) the management of composite devices for multimedia contents. Multimedia content caching and streaming are two essential solutions to the first issue (Paknikar et al., 2000). Web caching at a service provider site can largely increase the service provider's ability to support a large amount of service requesters. The streaming paradigm enables a media file be played at the service requester's site while it is being transferred over the Web; therefore, the burden of the service provider can be alleviated. Regarding the latter issue, at the service requester side, multimedia data is normally split over multiple devices with appropriate multimedia capabilities (Han, 2000).

Due to the rapid advancement of wireless information appliances, a Web service will gain more popularity if it is accessible from mobile devices in addition to normal Web browsers, such as wireless phones and Personal Digital Assistance (PDAs) (Pham, Schneider, & Goose, 2000). A Web service is thus considered to be device independent if it can be delivered to different devices (Han, 2000). Several techniques are designed to support the device independence. Based on an Extensible Markup Language (XML) and Resource Description Framework (RDF)-based framework, Composite Capability/Preference Profiles (CC/PP) are proposed to define device capabilities and user preferences so as to manage composite devices (CCPP). The combination of XML and Extensible Stylesheet Language (XSL)

is usually utilized to realize device independence (Kirda, 2001).

RELATED WORK

With the basic background, we proceed to review the related work on multimedia Web services.

Paknikar et al. (2000) define a client-side framework for the caching and streaming of Internet multimedia. The architecture consists of a number of caching proxy servers. A central controlling proxy server called a broker handles all of the initial interactions, and then transfers controls to its sibling proxy servers. This hierarchical structure provides scalability for the proxy servers. Their work also defines a layered replacement policy for caching scaleable encoded video objects. Paknikar et al. predicted that the Real Time Streaming Protocol (RTSP) will become the *de facto* standard for Internet Audio/Video (A/V) caching and streaming.

Pham, Schneider, & Goose (2000) define a *Small Screen/Composite Device (SS/CD)* architecture that supports small screen device-focused communication systems. The key component of the architecture is a Smart Gateway (SG) that distributes multimedia information to the most appropriate composite devices to ensure reliable performances. The critical component of SG is a set of algorithms associated with a Selection-Device-Assignment-Matrix. FieldWise (Fagrell, Forsberg, & Sanneblad, 2000) relies on a server engine to adapt multimedia responses according to the capabilities of the client devices and their network connections. However, these two projects do not adopt the most current Web techniques and standards, such as XML/XSL and CC/PP.

WebSplitter (Han, Perret, & Naghshineh, 2000) provides a unified XML framework supporting multi-device Web browsing. The framework defines an XML-based metadata policy file based on the CC/PP protocol to enable users to specify their access privilege groups. With WebSplitter, all Web pages are constructed as XML files, with pre-defined tags describing mappings to the corresponding access privileges. A proxy is then adopted to split a Web page to different devices. Corresponding XSL style sheets are attached to devices to transform the customized XML to the suitable device-understandable languages. MyXML (Kirda, 2001) is an XML/XSL-based template engine to solve the issue of device independence. The idea is to completely separate the content from its layout information. Similar to the WebSplitter, MyXML utilizes the XML/XSL combination to realize device independence. However, MyXML introduces a whole set of syntax elements that introduces a steep learning curve.

All these efforts concentrate on the client site to facilitate multimedia storage and streaming, to assist multimedia distribution, and to support device independence. At this moment, it appears that there still lacks a generic infrastructure for considering both the client and server sides of multimedia Web services. An additional limitation is that these methods may or may not integrate easily with the most current Web technologies and standards, since Web service is still an emerging paradigm. Here we seek to provide efficient support of delivery and management of multimedia Web services. In contrast with the previous approaches, we accomplish this objective by seamlessly incorporating the cutting-edge techniques of Web services and providing a SOAP-oriented component-based framework to support device-independent multimedia Web services.

SERVICE-ORIENTED COMPONENTIZATION MODEL

In this section, we will introduce our solution to support multimedia Web services. We will first discuss our idea of separation of control flow from data flow to utilize SOAP. Second, we will propose our enhancements to SOAP. Third, we

will propose enhancements to CC/PP protocol. Finally, we will propose our service-oriented multimedia componentization model.

Separation of Control Flow and Data Flow

As a core technique, SOAP is used to transport the content of Web services. However, SOAP was not originally designed to support multimedia Web services; therefore, its current version is not appropriate for streaming multimedia content. The reasons are multi-fold: (1) it is usually infeasible to put a large piece of multimedia content (e.g., a video clip) into one message. However, SOAP does not support message boxcarring and batching (SOAP); therefore, its current version cannot be used to transfer streamed multimedia content. (2) Using SOAP to transport data requires enormous network bandwidth due to its eXtensible Markup Language (XML) markup and protocol overhead (Werner, 2004). Therefore, its performance drawback hinders SOAP from transporting multimedia content that is usually associated with QoS requirements such as response time. (3) Current SOAP specification does not provide facility to define multimedia QoS requirements (Khan, 2002) and multimedia management information such as synchronization signals necessary for time dependent media. In summary, current SOAP is not suitable to transport massive amount of multimedia content.

We do not have to use SOAP to transport multimedia content. There are already a wealth of existing infrastructures, channels, and standards to support the transport of different multimedia content [e.g., MPEG-21 (MPEG-21), SMIL (SMIL), JPEG (JPEG), and HotMedia (HotMedia)]. We can still utilize these existing techniques to transport the content of media files. This, however, still leaves the transport of the multimedia control information?

Our solution is to separate the control information from the content information. The control information may include: (1) meta data that depict the synchronization relationship between multiple media files, (2) QoS requirements, and (3) other control information such as the service provider. Thus transporting a multimedia Web service includes the delivery of both data flow and control flow. The essential advantage is that different transport protocols can be employed to deliver either data flow or control flow, that is, SOAP can be used to transport the control information while traditional multimedia channels can be used to transport the multimedia content in SOAP-specific environment.

This separation of control flow and data flow promises several merits. First, the separation solves the dilemma of transferring multimedia applications into multimedia Web services. Control information will be delivered by SOAP so that different multimedia Web services can interoperate with each other, while multimedia content information can be delivered by traditional multimedia protocols to achieve acceptable performance. Second, the separation provides a loose coupling between control information and content information; thus, the content information can be reused for different Web services. Third, the separation facilitates the tracking and logging of control information so that we can achieve better management and monitoring of multimedia Web services.

Enhancements of SOAP to Support Multimedia Web Services

SOAP has been considered to be a *de facto* communication standard to deliver Web services; however, it was not originally designed particularly to deliver multimedia Web services. First, the nature of multimedia data (Khan, 2002) requires the fact that the transportation of the multimedia information normally has to meet some QoS requirements, such as the synchronization within and among different multimedia data streams or real-time delivery. Consequently, a message

containing multimedia data should carry over its QoS requirements as the guidance for Internet transportation. Nevertheless, the original SOAP specifications do not support this feature. Therefore, we believe that enhancements to SOAP are compulsory in order to facilitate multimedia Web services.

Second, for simplicity, SOAP does not support boxcarring and batching of messages; also it is a one-way protocol (SOAP). There are cases where it is difficult to embed a large multimedia file into one SOAP message. On the contrary, it may be more practical to load a big chunk of information into multiple SOAP messages. Of course there should be a way to identify the relationships among these related SOAP messages. In exceptional cases, some of the media files may not even be suitable to be transferred in SOAP messages; for example, a multimedia segment may need to be transferred in one file with specific file extension. Therefore, special attributes should be provided to identify all of these situations. Although, as we pointed out previously, multimedia information may be transported via channels other than SOAP, nevertheless, for completeness, we still believe that SOAP should be extended with mechanisms to transport multimedia information.

Our enhancements to SOAP can be therefore categorized into two sets: message batching specifications and multimedia QoS specifications. Two sets of attributes are introduced as summarized in *Table 1*; and each of them will be discussed in detail in this section. We first propose a simple workaround of boxcarring and batching of messages to bolster large-scaled multimedia data transportation. A service provider can divide a big message into multiple smaller messages; and these smaller messages altogether conceptually constitute a message box. The service provider can then send these messages to the service requester asynchronously.

As illustrated in the first part of *Figure 1*, several attributes are defined to support this ability: *SenderURL*, *id*, *index*, and *total*. *SenderURL* is defined to specify the unique address of the service provider, so that the service requester knows where to fetch further information if so desired; id is defined to uniquely identify the message box in the domain of the service provider; index is defined to identify the index of the message in the corresponding message box; and total is defined to identify the total number of messages in the message box. These attributes can be utilized to uniquely identify a SOAP message and

Figure 1. A piece of SOAP header message

```
<SOAP-ENV:Envelope
….
      SOAP-ENV:MustSendBack=”RequestId0001, Profile001”
      SOAP-ENV:id=”Msg00001”
      SOAP-ENV:index=”1”
      SOAP-ENV:total=”5”/>
      <SOAP-ENV:Body>
         <m:Metadata>
              <component>Msg00001”</component>
              <component>Msg00002”</component>

              <component>Msg00003”</component>
              <component>Msg00004”</component>
              <component>Msg00005”</component>
         </m:Metadata>
       </SOAP-ENV:Body>
….
</SOAP-ENV:Envelope>
```

Table 1. SOAP attributes introduced

Attribute	*Definition*
SenderURL	Unique address of the service provider
id	Uniquely identify the message
index	The index in message box
total	The total number of messages in message box
MustSendBack	Required to be sent back without changes
Metadata	Specify the structure of the message box
component	Specify the messages in the message box

Attribute	*Definition*
reliable	Reliable transportation
realTime	Real-time transportation
unicast	Unicast to specific node
multicast	Multicast to multiple nodes
secure	Secure in the transportation

the relationships between messages. In addition, a global attribute *MustSendBack* is introduced, which can be utilized to specify that the information is required to be sent back to the service provider without any changes. For example, the service requester's profile needs to be sent back along with the result for the proper multimedia distribution.

Employing the metaphor of the envelope in postal delivery, we insert these identification information of the multimedia content into the corresponding SOAP envelope as the first part. Furthermore, the first message in the message box should contain the metadata in its body block, which identifies the structure of messages contained in the message box. Attributes Metadata and component are defined to specify the metadata information. Let us take a piece of a SOAP message as an example, as shown in *Figure 1.*

This SOAP message is the response message to the request "RequestId0001". This request id "RequestId0001" and user profile information "Profile001" must be sent back. We can see that the message is identified by its unique id "Msg00001"; and it is the first of the total five messages in the message box. Metadata records the message ids of all five messages: "Msg0001," "Msg0002," "Msg0003," "Msg0004," and "Msg0005." Therefore, the requester that receives this message would be aware of the remaining four messages, then could schedule to pre-fetch them with specified ids if so desired.

Second, we introduce five global attributes to the SOAP definition in order to enable a SOAP message to carry on its QoS requirements: *reliable*, *realTime*, *unicast*, *multicast*, and *secure*. These five attributes are summarized in the second part of *Table 1.* For an attribute to be set, its keyword-value pair must be present in this envelope block. The *reliable* attribute can be used to indicate whether the SOAP message requires reliable transportation or not. The value of the *reliable* attribute is either "1" or "0." The absence of the *reliable* attribute is semantically equivalent to its presence with a value "0." If a header element is tagged with a *reliable* attribute with a value of

"1," the recipient of that header entry either MUST find a reliable transportation protocol, or MUST fail processing the message.

The *realTime* attribute can be used to indicate whether the SOAP message requires real-time transportation or not. The value of the *realTime* attribute is either "1" or "0." The absence of the *realTime* attribute is semantically equivalent to its presence with a value "0." If a header element is tagged with a *realTime* attribute with a value of "1," the recipient of that header entry either MUST find enough resources to transfer the message right away, or MUST fail processing the message.

The *unicast* attribute can be used to indicate whether the SOAP message is to be unicasted to a specific end point. The value of the *unicast* attribute is either "1" or "0." The absence of the *unicast* attribute is semantically equivalent to its presence with a value "0." If a header element is tagged with a *unicast* attribute with a value of "1," the recipient of that header entry either MUST find a available path to the specified end point node, or MUST fail processing the message.

The *multicast* attribute can be used to indicate whether the SOAP message is to be multicast to multiple users. The value of the multicast attribute is either "1" or "0." The absence of the *multicast* attribute is semantically equivalent to its presence with a value "0." If a header element is tagged with a *multicast* attribute with a value of "1," the recipient of that header entry either MUST find available paths to all specified end-point users, or MUST fail processing the message.

The *secure* attribute can be used to indicate whether the SOAP message has to be kept secure in the process of transportation. The value of the *secure* attribute is either "1" or "0." The absence of the *secure* attribute is semantically equivalent to its presence with a value "0." If a header element is tagged with a secure attribute with a value of "1," the recipient of that header entry either MUST find secure paths to the destination, or MUST fail processing the message.

Figure 2 is another example of a piece of a SOAP message containing some multimedia information. This message needs to be transferred reliably, real-time, and only transferred to one specific end user. This enhancement provides a way for a SOAP message to delineate its QoS requirements, which can be utilized as a guidance of Internet transportation tools to select different paths and transportation protocols so as to increase performance.

Enhancement of CC/PP Supporting Multimedia Web Services

Multimedia Web services normally need to split multimedia data over multiple devices with appropriate multimedia capabilities (Han, Perret, & Naghshineh, 2000). CC/PP specifies an XML and RDF-based framework to help define device capabilities and user preferences so as to manage composite devices (CCPP). We adopt CC/PP to support multimedia Web services for two reasons. One is that CC/PP was designed particularly for describing device capabilities and user preferences. The other one is that CC/PP is built on top of popular XML technology. Utilizing the CC/PP format, every user can create his/her own profile that declares the list of the user's available re-

Figure 2. A second piece of SOAP message example

```
    <SOAP-ENV:Envelope  xmlns:SOAP-
ENV="http://schemas.xmlsoap.org/soap/
envelope/"
          SOAP-ENV:encodingStyle="http://
schemas.xmlsoap.org/soap/encoding/"/>

          SOAP-ENV:reliable="1"
          SOAP-ENV:realTime="1"
          SOAP-ENV:unicast="1"
    …
          <SOAP-ENV:Body>
    …
          </SOAP-ENV:Body>
    …
</SOAP-ENV:Envelope>
```

sources (devices) and preferences about how each resource would be utilized. However, we found that CC/PP definition is too rigid for device capabilities. For instance, one device may be capable of accepting one kind of media information with some simple transformations; and CC/PP does not support this specification. Therefore, we extend CC/PP by enabling transformation description to be added to devices.

With our extension, in a user's profile, every resource is declared as a separate item, such as a WAP phone. Each item comprises of two parts: one is the declaration of the hardware, and the other one is the declaration of the service, or more directly multimedia resource files that the device can receive. *Figure 3* is an example of a piece of an extended-CC/PP profile for a WAP phone.

As shown, the declaration of this WAP phone contains two parts. The first part defines the hardware, such as the device name, the size of the screen, resolution, and etcetera. The second part defines its acceptable multimedia information, such as image, text, WML, and etcetera. However, it can be noticed that the normal HTML code cannot be directly retrieved on this WAP phone. A stylesheet named HTML2WML.xsl needs to be used to transform the receiving HTML code to the WAP-enabled WML code. With this extension, CC/PP protocol becomes more powerful of describing device capabilities.

Proposed Model Supporting Multimedia Web Services

As Roy and Ramanujan (2001) summarized, a three-role model is broadly adopted in the field of Web services: service providers, service brokers, and service requesters. From the perspective of a service requester, there are two key questions: how to find a demanded service, and how to get the service effectively. This paper focuses on the solution to the second question; therefore, service brokers and related issues will not be discussed. The starting point of this paper is that, with the help of service brokers, service requesters have already located the service providers that offer the requested services. Meanwhile, as high-speed local area networks (LANs) have been deployed extensively over the world (Paknikar et al., 2000), users on such LANs normally access Internet through a proxy server that provides caching facility. Therefore, we make another assumption on such a concept, that users always invoke Web services through their proxy server. In addition, due to the fact that service requesters are normally connected from service providers by Internet, it

Figure 3. A piece of cc/pp profile example

```
<ccpp:component>
    <rdf:Description rdf:about="TerminalHardware">
        <rdf:type rdf:resource="HardwarePlatform"/>
        <DeviceName>Nokia-3360</DeviceName>
        <screen>30X23mm</screen>
        <display>101X52Pixels</display>
        <PixelStretch>1.24</PixelStretch>
    </rdf:Description>
</ccpp:component>
…
<ccpp:component>
    <rdf:Description rdf:about="Services">
        <rdf:type rdf:resource="SupportedServices"/>
        <rdf:li>HTML</rdf:li>
        <rdf:transform>HTML2WML.xsl</rdf:transform>
    </rdf:Description>
</ccpp:component>
```

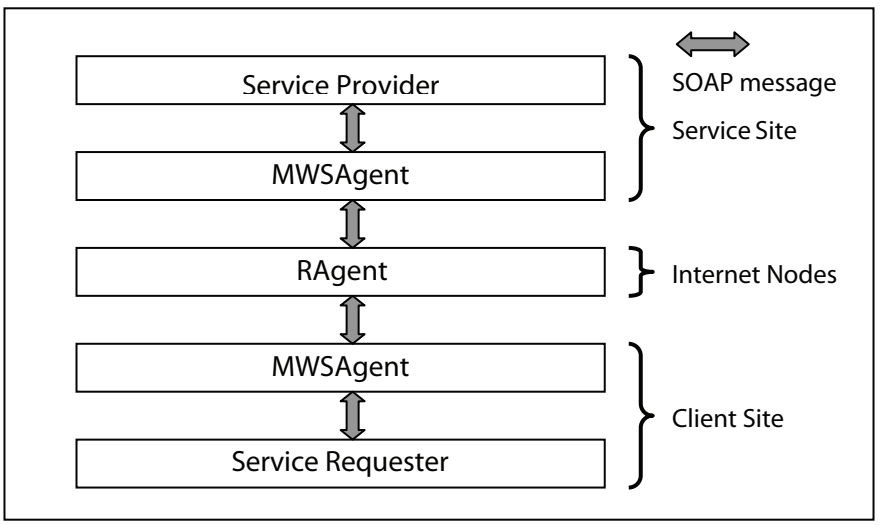
is reasonable to assume that a message has to pass through multiple networks between them. Based on these three assumptions, the problem can be refined as: how to effectively and efficiently support multiple service requesters, who rely on the same proxy server and request the same Web service from the same service provider that is several networks away.

Here we propose a component-based framework as a solution. As illustrated in *Figure 4*, three types of intelligent agents are introduced as the main components of the framework; and SOAP is adopted as the service transportation protocol. The first agent introduced is the Multimedia Web Service Server Agent (MWSSAgent) that locates at service providers; the second one is the Multimedia Web Service Agent (MWSAgent) that locates at proxy servers on the clients' LAN; and the third one is the Routing Agent (Ragent) that locates at intermediate network nodes. *Figure 4* as well implicates the information path of a multimedia Web service on the basis of this framework. When a service requester submits the request, the MWSSAgent handles the request, invokes the corresponding service backend to get the results, equips return SOAP messages, and sends them to the Internet. The RAgent reads the envelopes of the receiving SOAP messages, analyzes their QoS requirements, and selects the appropriate protocols to send to the next proper RAgent or the proxy server. The MWSAgent finally receives the SOAP messages, propagates to the original requester, and distributes the multimedia streams to appropriate media devices, such as Web browsers, audio systems, PDAs, and etcetera. We will discuss each of the three agents in detail in the following sections.

Multimedia Web Services Server Agent (MWSSAgent)

A multimedia Web service may generate a complex result including multiple multimedia files. Disregarding the fact that SOAP currently can only bind to HTTP while HTTP was not designed for streaming media data (Paknikar et al., 2000), it is not efficient and practical to encapsulate all information in one SOAP response message to pass through the Internet, especially when the information may get lost in the process of transportation. With our enhancements to SOAP described in the previous section, a big piece of information

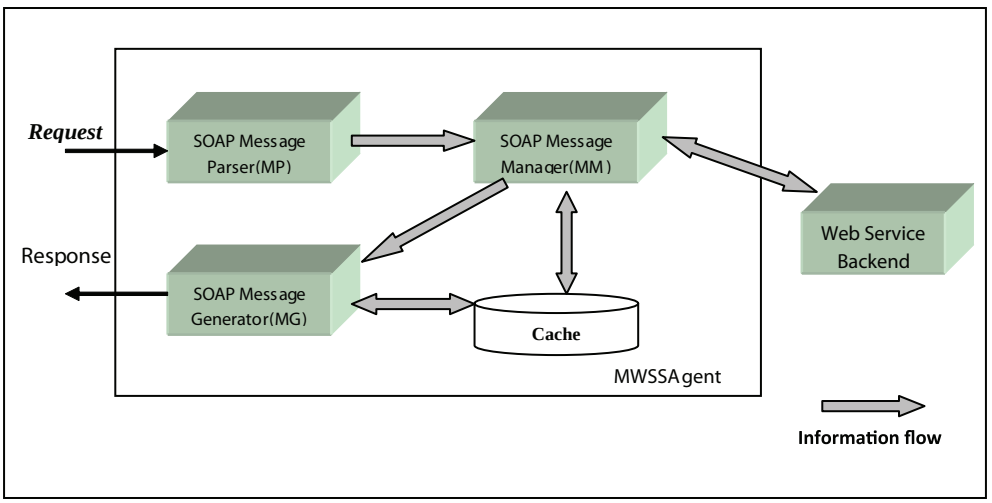
Figure 4. A component-based framework for multimedia Web services



can be separated into multiple interrelated SOAP messages, with the first message containing the metadata and specifying the identification of each message. The metadata is completely separated from the real content. Here the metadata includes the structure of the set of the sub-messages, and the relationships between messages. It is one of the main responsibilities of the WMSSAgent to generate the SOAP response message box from the result provided by the service provider. To avoid aggravating Internet traffic, the response philosophy that the MWSSAgent adopts is a lazy-driven norm, that is, it does not always send back to the service requester all of the messages at the same time; some messages may stay on the MWSSAgent until they are requested particularly. Therefore, the caching facility is necessary to support this philosophy. In addition, published Web service providers normally expect large amount of requests, thus managing and reusing cached messages become essential.

Therefore, we propose the MWSSAgent as an intelligent agent on the service provider site to facilitate multimedia services. Its architecture is illustrated in *Figure 5*, together with the interactions among its components. The architecture contains three functional components — a SOAP message parser (MP), a SOAP message generator (MG), a SOAP message manager (MM), with a local cache. MP is in charge of parsing incoming SOAP request messages, interpreting the requests and forwarding them to MM. MM first checks the cache to see whether the result SOAP messages have already been generated and stored. If the results exist, the MM will send back the results through the MG — the MG needs to generate the corresponding return envelopes based on the requests. If the results are not found, the MM will invoke the service backend for the service. The MG will then generate the full set of SOAP response messages and store to the local cache, before sending back the first response message that includes the metadata. All of the SOAP response messages will be cached on the local disk under the control of the MM. When a SOAP request message arrives, the MP will verify whether the request is the first request for a service, or a subsequent request for a specific response message of a particular service. When the MG generates the set of response messages, all messages will be uniquely numbered for identification purposes. Owing to the fact that a service requester may invoke a specific SOAP message later on, the full set of SOAP messages corresponding to a Web

Figure 5. MWSSAgent architecture



service will either be all stored in the cache of the MWSSAgent, or fully swapped out when storage limitation is encountered. Here we adopt the caching and streaming algorithms introduced in Paknikar et al. (2000).

Multimedia Web Services Agent (MWSAgent)

We propose a MWSAgent as an intelligent agent on the proxy server at the client site to support multimedia Web services. The component-based architecture is illustrated in *Figure 6*. The MWSAgent comprises of four functional components: service broker, a user profile manager, a service delivery manager, and a multimedia manager. In addition, there are three caches contained in a MWSAgent: one for user profiles, one for request-profile mapping, and one for multimedia information.

User Profile Manager (UPM)

The User Profile Manager (UPM) registers user profiles and stores them in the user profiles cache. Adopting our extended CC/PP protocol as discussed in the previous section, a user profile declares the list of the user's available resources and the user's preferences about how each resource would be utilized. UPM assigns a unique id to every registered profile, so that one user can possess multiple profiles for the purposes of different Web services. When a user requests a Web service, she is required to specify which profile she wishes to use to receive the service. The service broker will assign a unique id to each request; and the UPM will store the 2-tuple (request id, profile id) in the cache of request-profile mapping. This 2-tuple will also be included in the SOAP message of the service request as *MustSendBack*. Therefore, when the response comes in, the UPM will use the sent-back (request id, profile id) pair to pick up the stored corresponding profile from the profile cache. Users are allowed to change their profiles, add new profiles, or delete some profiles. The granularity of the definition of a profile depends on the power of CC/PP and RDF vocabulary.

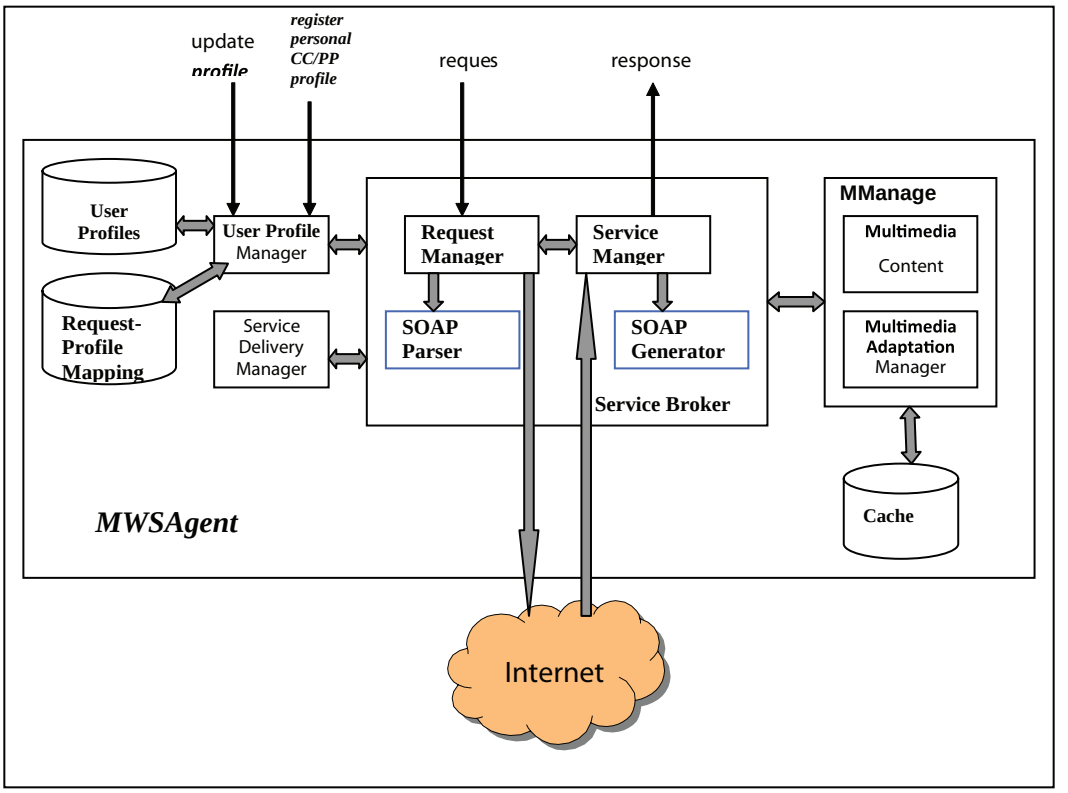
Multimedia Manager (MManager)

The Multimedia Manager (MManager) contains two sub-components: the Multimedia Content Manager (MCM) and the Multimedia Adaptation Manager (MAM). The MCM manages the caching of multimedia SOAP response messages on the local disk of the proxy server. The MAM handles media adaptation if necessary, in order to achieve device independence. In some cases when the destination device is not capable of handling requested multimedia contents, the MAM may convert the media accordingly. In some other cases, the destination device may announce that it is able to receive some media contents after some adaptations are performed. For instance, a WAP phone could accept HTML code; however, it needs to be transformed to WML code. In accordance with our extension to the CC/PP, when a user specifies her profile, she can stipulate the transformation algorithm she prefers.

Service Broker

The service broker consists of four sub-components: a request manager, a service manager, a SOAP generator, and a SOAP parser, as shown in *Figure 6*. The SOAP generator generates SOAP request messages; the SOAP parser parses incoming SOAP response messages; and the request manager is a request broker of user requests for Web services. A user may request a Web service that another user from the same LAN has previously requested. Therefore, the request manager will first check the local disk through the MCM. If the result is a hit, the request manager will pass the control to the service manager to send back the information. As a consequence, not only the remote service broker and the service provider will have less traffic, but also the response time

Figure 6. MWSAgent architecture



may be largely shortened. Otherwise, the request manager will assign a unique id to the request, ask UPM to store to the request-profile mapping, and then call the SOAP generator to generate a SOAP request message and send it to the service provider over the Internet. The unique request id and the user profile id will be marked as *MustSendBack* and be sent together with the SOAP request.

The service manager is invoked when a SOAP response comes back from Internet. It will first call the SOAP parser to analyze the content of the result message. If the result contains multimedia information not coming together with the first message, the service manager will schedule to pre-fetch the corresponding rest of media files to enhance the streaming of the media data. All the messages will be sent to the MCM to store in the local disk before sending back to the original requester, so as to achieve caching at the proxy server level. As a result, the service provider might not be even online all the time but its content can still be available at the cache of the MWSAgent. For persistent multimedia data (e.g., a movie), this cached data may be immediately utilized. For time-varying data, however, a mechanism for determining if the data is stale will have to be employed. Consider scalable encoded or layered video information for example: such objects have a "base" layer containing essential information, and one or more "enhanced" layers containing higher level information (Paknikar et al., 2000). The service manager will try to download the sub-nodes following the layers. It will attempt to download the lower layers before downloading higher and more enhanced layers. Meanwhile, the service manager will work with the UPM to launch the corresponding user profile. And then the control will be passed to SDM to deliver the result to the original service requester.

Service Delivery Manager (SDM)

The Service Delivery Manager (SDM) decides the priority and the order of the services to be sent back to the requesters, and decides how to split the returning information to appropriate devices if there are multiple devices. A set of criteria is kept at the SDM to be utilized to optimize the order of the delivery, such as request time, request priority, current available resources, and etcetera. The details of the algorithms of the selection and adaptation will not be discussed in this paper.

Route Agent (RAgent)

SOAP messages are adopted to transfer requests and responses through the Internet between a MWSSAgent and a MWSAgent. A SOAP message in this paper contains multimedia information that normally possesses QoS requirements. As discussed in the previous section, a SOAP message may have to travel through several heterogeneous intermediate networks before it finally reaches the MWSAgent. Each of these heterogeneous networks could support multiple protocols and the flexibility of selecting protocols dynamically (Banchs et al., 1998). Therefore, in order to satisfy the performance requirements of the multimedia SOAP message to be transported, each network may need to select the most appropriate network protocol. For example, a NACK-reliable multicast protocol (Floyd et al., 1997) should be adopted on an ATM network in order to increase performance. We propose the RAgent to achieve this goal. A RAgent resides at an intermediate network node on the Internet. Generally we suggest that each active network install a RAgent as the high-level director of the network routing. For a legacy system network that does not install RAgent, SOAP messages can be passed without looking at the information it carries.

This paper focuses on the selection of protocols to increase the performance of network transport rather than the selection of path-routing algorithms. The architecture of a RAgent is composed of four sub-components, as illustrated in *Figure 7*. The SOAP parser is responsible for parsing incoming SOAP messages, determining the QoS requirements, and passing them to the Protocol Manager (PM). The PM receives QoS requirements, searches the Protocol Pool (PP), and locate the appropriate protocol to use. The PP is the interface encapsulating all protocols registered in the corresponding network. The Protocol Registration Manager (PRM) handles the registration of new protocols to the network or removal of some protocols. Under the PP are two sub-components: a cache of the registered protocols, and an abstract matrix that records the protocols and their QoS characteristics, such as reliability, real-time, security. The PRM maintains the abstract matrix when a new protocol is registered to the PP or removed from the PP.

When the PRM registers a new protocol to the protocol pool, it not only records its QoS properties, but also its cost. Typically, the cost here refers to the efficiency of the protocol. The more reliable a protocol is, the more costly it will be. For example, TCP is more costly than UDP because of the acknowledgment requirement. *Table 2* is an example of the abstract protocol matrix. Three protocols are registered into the protocol pool: TCP, UDP, and NACKRMP. TCP and NACKRMP are reliable protocols, while NACKRMP is a real-time protocol. TCP and UDP can be used for unicasting to a single node, while TCP, UDP, and NACKRMP can all be adopted for the purpose of multicasting. From *Table 2* we can also conclude that TCP is more costly than UDP; and NACKRMP costs the most.

PERFORMANCE EVALUATION

The goal of our experiments is to design and conduct a simulation to evaluate the performance of our framework in a typical multimedia Web service application. The application we selected

Figure 7. RAgent architecture

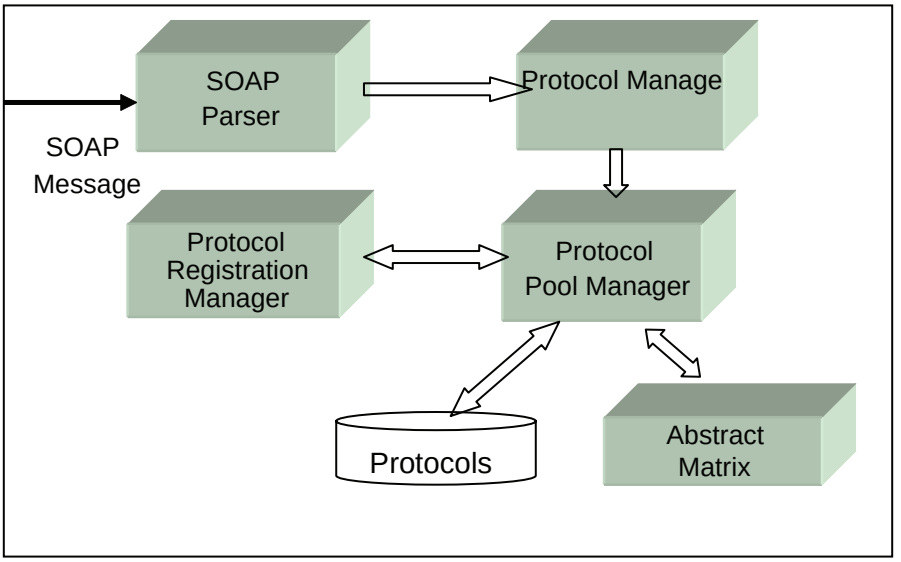


Table 2. Example of abstract protocol matrix

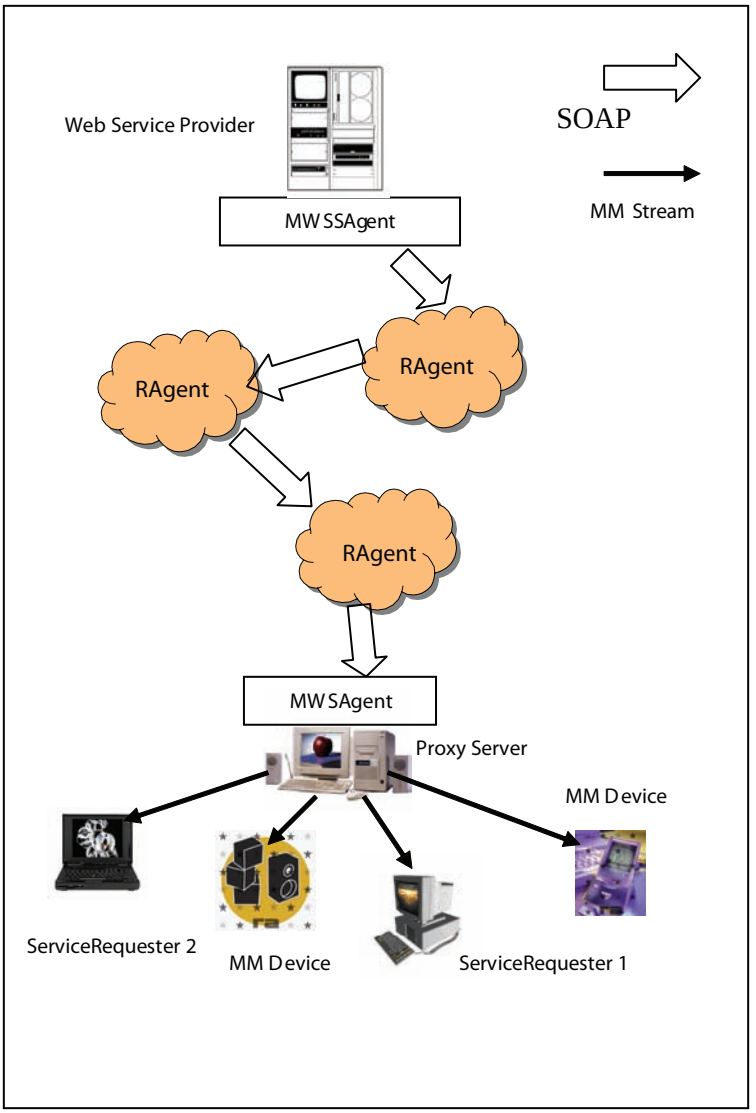
Protocol	*TCP*	*UDP*	*NACKRMP*
Reliability	**X**		**X**
Security			
Real-time		**X**	**X**
Uni-cast	**X**	**X**	
Multi-cast	**X**	**X**	
Cost	**4**	**1**	**5**

to implement is a distance-learning environment: multiple students request multimedia course information from servers. In such an application, one may assume that the students will be congregated at a set of proxy servers, and will be accessing a common set of material. Hence, we want to vary the number of service requesters and proxy servers, holding the other system components constant. To reflect this requirement, we design the experiment based on the following assumptions. First, we assume that there is only one service provided by the system, and there is only one server machine that provides the service. Second, we assume that multiple students who reside on the same LAN request the service through one common proxy server. Third, we assume that there are three networks between the server and the proxy server: Ethernet and ATM and then Ethernet. A message from the server machine has to first pass through these three networks in the order specified before it hits the proxy server. Ethernet has TCP and is NACK RMP (Floyd et al., 1997) registered; and the ATM network has UDP and is NACK RMP registered. Suppose we require real-time QoS multimedia transportation. Fourth, we assume that the transportation time between a service requester and its corresponding proxy server is ignored due to the fact that they reside on the same LAN. Therefore, the problem is simplified and refined as shown in

Figure 8. One server machine stores all of the course information and serves as the service provider; and multiple students request the same course information from this server machine as service requesters. Each student may possess one computer and some multimedia devices, such as speakers or cell phones. All students reside on the same LAN and communicate to the service provider through a proxy server machine.

We set up the experimental environment as shown in *Figure 8*. The multimedia Web service is implemented as a normal J2EE-compatible Web application on the JRun application server (JRun). One MWSSAgent is installed on the server; and one MWSAgent is installed on the proxy server. Each of the three networks has one machine that applies RAgent. All of the three types of agents are implemented in Java. Both MWSSAgent and MWSAgent are also implemented as J2EE-compatible server on the JRun application server (JRun). The SOAP parsers and generators on all three agents are implemented by modifying the

Figure 8. An example applying framework



open source Apache Axis system (Axis). According to our experimental setup, the two RAgents in the Ethernet will choose TCP, and the RAgent in the ATM network will choose UDP to increase the performance.

We design the experiment such that each client would create threads and then communicate with the server, requesting multimedia services. We perform tests on the server machine applying a

MWSSAgent and without a MWSSAgent. Moreover, we performed the same set of tests on the service with one SOAP message (980KB) and two SOAP messages (980KB each). For our experimentation we performed very low contention (1 client), low contention (2,4 clients), moderately low contention (6,8 clients), moderately high contention (10,12,14 clients), and high contention (16,18 clients). For each client, we record the response time, from the time the thread starts the request to the time the entire multimedia information is received. Then the response time from all clients were averaged. The results performed on the four sets of situations are shown in *Figure 9*. Each point in the figure shows the averaged response time measured in seconds when there are specific numbers of clients requesting the service. From the figure, it shows that the system applying MWSSAgent outperforms the system without it. The higher number of clients the system supports, the higher the performance the MWSSAgent exhibits. *Figure 9* also shows that the more clients a service provider needs to support, the higher efficiency our framework exhibits. Only when there is one client who requests a service, the system applying the MWSSAgent takes longer because of the extra operation time spent on the MWSSAgent. However, the benefits justify this extra work as long as there are multiple requesters.

To test the effectiveness and efficiency of applying the MWSAgent, we relax one assumption from the first experiment, so that there are two proxy servers in the experimental system, and clients are averagely distributed behind these two proxy servers. We apply the MWSAgent on each of the proxy servers. The results performed on the same four sets of situations are illustrated in *Figure 10*. The approach to measure the results is the same as that of the first experiment. The figure shows that the system applying MWSAgent outperforms the system without it. *Figure 10* shows

Figure 9. One-proxy performance

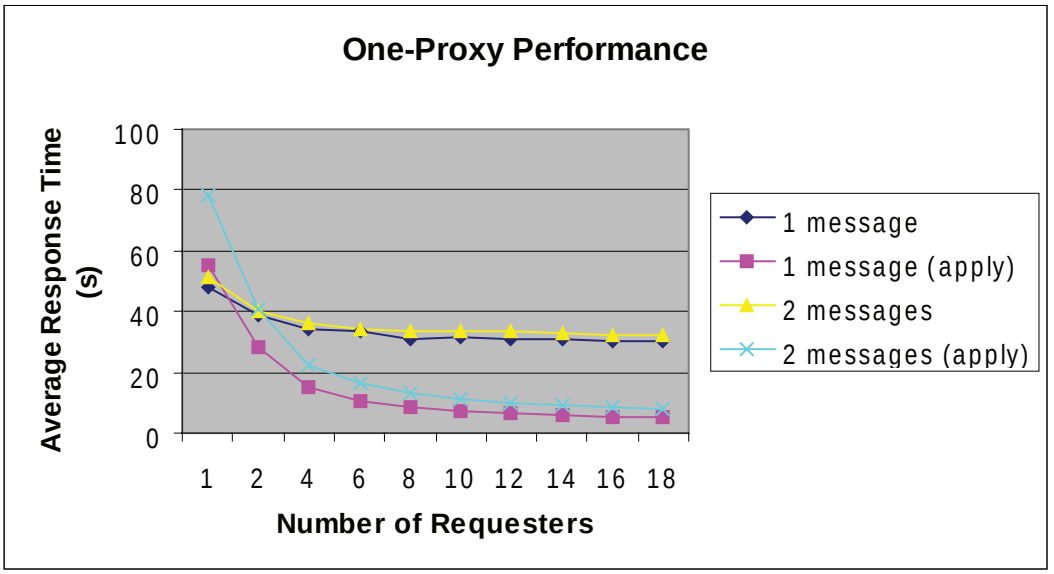
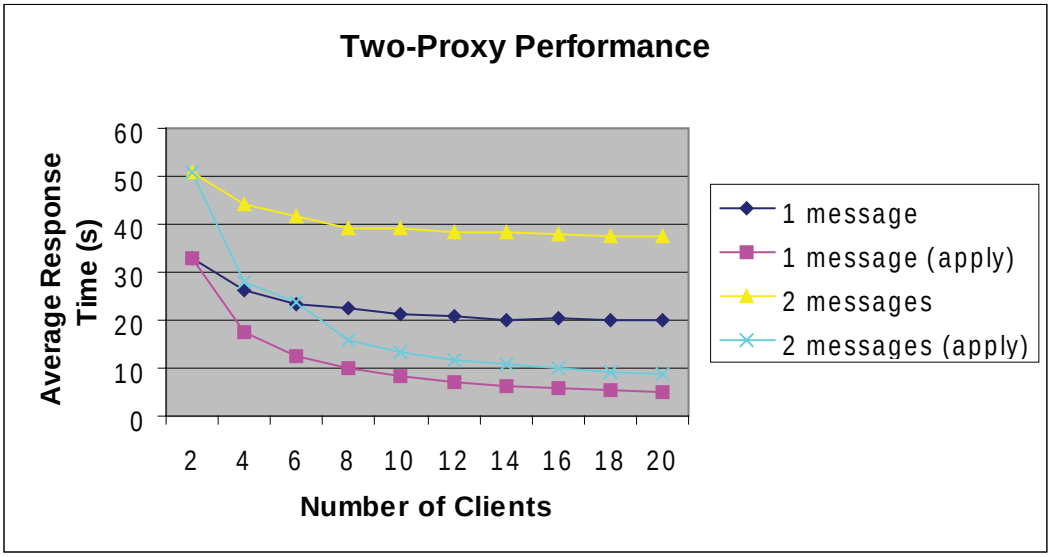


Figure 10. Two-proxy performance



that the more clients sharing one proxy server, the higher efficiency the MWSAgent exhibits. It also shows that when there is only one client behind of a proxy server, the system applying the MWSAgent takes longer because of the extra operation time overhead spent on the MWSAgent. However, again, the benefits justify this extra work as long as there are multiple requesters.

These two experiments show that our framework exhibits the distinct advantage of facilitating a multimedia Web service provider to support larger amounts of Internet clients, and shortening the average response time for service requesters.

ASSESSMENTS, INNOVATIONS AND FUTURE WORK

We present in this paper a service-oriented componentization model to support device-independent multimedia Web services. In the current infrastructure, we adopt caching and replacement algorithms introduced in Paknikar et al. (2000) for both the MWSAgent and the MWSSAgent. Since these two agents serve different purposes, using the same set of caching and replacement algorithms may not be most efficient. In addition, this paper concentrates on protocol selection and dynamic binding, based on the QoS requirements carried by SOAP messages, on networks to support multimedia QoS requirements. To guarantee multimedia QoS requests, efficient routing algorithms are inevitable. Many challenging issues remain in the realm of multimedia Web service. The framework proposed in this paper is based on a much simplified problem domain.

Despite these limitations that could be improved upon or resolved by further work, our model extends research on multimedia Web services in several ways. First, the SOAP enhancements provide a simple way to improve the ability and flexibility of the *ad hoc* standard SOAP protocol to serve for multimedia Web services, by supporting batch facilities and QoS requirements. Second, the CC/PP enhancements increase the flexibility of the system to manage multi-devices and ensure device independency. Third, three types of intelligent agents are synergistically integrated to form a framework to

support efficient service-oriented multimedia Web services. This model also facilitates caching and streaming of multimedia transport. In addition, our model seamlessly incorporates cutting-edge technologies relating to Web services: SOAP, XML/XSL, and the CC/PP. The result of this research can be applied to the software industry and serves as an architectural design to construct multimedia Web service applications.

We intend to continue our research work in the following directions. First, we will explore efficient caching, replacement, and pre-fetching algorithms so as to improve QoS performance. Second, we will attempt to bind SOAP to other more multimedia-oriented transportation protocols, such as Real Time Streaming Protocol (RTSP). Third, we will investigate QoS routing algorithms on routers to support SOAP QoS requirements. Fourth, we will pursue a formal description language to facilitate protocols to be published on the Web and dynamically registered to networks. Finally, we intend to implement a mechanism to monitor multimedia QoS performances over different networks.

REFERENCES

Axis. (2004). *Apache AXIS Project, SOAP Protocol Implementation*. Retrieved from http://ws.apache.org/axis.

Banchs, A., Effelsberg, W., Tschudin, C., & Turau, V. (1998). Multicasting multimedia streams with active networks. In *Proceedings of the 23rd Annual Conference on Local Computer Networks* (pp. 150-154), October 11-14, 1998, Boston, MA, USA.

Benatallah, B., Sheng, Q.Z., & Ngu, A.H.H. (2002). Declarative composition and peer-to-peer provisioning of dynamic Web services. In *Proceedings of the 18th International Conference on Data Engineering* (ICDE'02) (pp. 297-308). February 26-March 01, 2002, San Jose, CA, USA.

CCPP. (2001, March 15). *W3C Composite Capability/Preference Profile*. W3C Working Draft.

Fagrell, H., Forsberg, K., & Sanneblad, J. (2000). FieldWise: A mobile knowledge management architecture. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'00)* (pp. 211-220). Philadelphia, PA, USA.

Ferris, C., & Farrell, J. (2003). What Are Web services? *Communications of the ACM*, *46* (6), 31.

Floyd, S., Jacobson, V., Liu, C.G., McCanne, S., & Zhang, L. (1997). A reliable multicast framework for light-weight sessions and application level framing. *IEEE/ACM Transactions on Networking, 5* (6), 784-803.

Han, R., Perret, V., & Naghshineh, M. (2000). WebSplitter: A unified XML framework for multi-device collaborative Web browsing. In *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work* (CSCW'00) (pp. 221-230). Philadelphia, PA, USA.

HotMedia. (2004). Retrieved from *http://www-306.ibm.com/software/awdtools/hotmedia*

JPEG. (2004). Retrieved from *http://www.jpeg.org.*

JRun. (2004). Retrieved from *http://www.macromedia.com/software/jrun*

Khan, M.F., Ghafoor, H., & Paul. R. (2002). QoS-based synchronization of multimedia document streams. In *Proceedings of IEEE 4th International Symposium on Multimedia Software Engineering* (MSE'02) (pp. 320-327). December 11-13, 2002, Newport Beach, CA, USA.

Kirda, E. (2001). Web engineering device independent Web services. In *Proceedings of the 23rd International Conference on Software Engineering* (ICSE'01) (pp. 795-796). Toronto, Ontario, Canada,.

MPEG-21. (2004). Retrieved from *http://xml.coverpages.org/ni2002-08-26-b.html*

NORTEL. (2004). Retrieved from *http://www.nortelnetworks.com/products/01/mcs52/collateral/nn105360-091103.pdf*

Paknikar, S., Kankanhalli, M.S., Ramakrishnan, K.R., Srinivasan, S.H., & Ngoh, L.H. (2000). A caching and streaming framework for multimedia. In *Proceedings of the 8th ACM International Conference on Multimedia* (pp. 13-20). Marina del Rey, CA, USA.

Pham, T., Schneider, G., & Goose, S. (2000). A situated computing framework for mobile and ubiquitous multimedia access using small screen and composite devices. In *Proceedings of the 8th ACM International Conference on Multimedia* (pp. 323-331). Marina del Rey, CA, USA.

Roy, J., & Ramanujan, A. (2001, November). Understanding Web services. *IEEE IT Professional*, 69-73.

SMIL. (2004). *W3C Synchronized Multimedia Activity Statement.* Retrieved from *http://www.w3c.org/AudioVideo/Act ivity.html*

SOAP. (2004). *Simple Object Access Protocol (SOAP) 1.1.*

UDDI. (2004). Retrieved from *http://www. uddi.org*

Werner, C., Buschmann, C., & Fischer, S. (2004). Compressing SOAP messages by using differential encoding. In *Proceedings of IEEE International Conference on Web Services* (ICWS'04) (pp. 540-547). July 6-9, 2004, San Diego, CA, USA.

WSDL. (2004). Retrieved from *http://www.w3.org/TR/wsdl*

This work was previously published in International Journal of Web Services Research, Vol. 2, No. 1, edited by L. Zhang, pp. 54-76, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section 3

Tools and Technologies

This section presents extensive coverage of the interaction between multimedia and the various tools and technologies that researchers, practitioners, and students alike can implement in their daily lives. These chapters provide an in-depth analysis of mobile multimedia, while also providing insight into new and upcoming technologies, theories, and instruments that will soon be commonplace. Within these rigorously researched chapters, readers are presented with countless examples of the tools that facilitate the transmission of multimedia data. In addition, the successful implementation and resulting impact of these various tools and technologies are discussed within this collection of chapters.

Chapter 3.1
Multimedia Content Representation Technologies

Ali R. Hurson
The Pennsylvania State University, USA

Bo Yang
The Pennsylvania State University, USA

INTRODUCTION

Multimedia: Promises and Challenges

In recent years, the rapid expansion of multimedia applications, partly due to the exponential growth of the Internet, has proliferated over the daily life of Internet users. Consequently, research on multimedia technologies is of increasing importance in computer society. In contrast with traditional text-based systems, multimedia applications usually incorporate much more powerful descriptions of human thought – video, audio and images (Auffret, Foote, Li & Shahraray, 1999). Moreover, the large collections of data in multimedia systems make it possible to resolve more complex data operations, such as imprecise query or content-based retrieval. For instance, image database systems may accept an example picture and return the most similar images of the example (Cox, Miller & Minka, 2000, Huang, Chang & Huang, 2003). However, the conveniences of multimedia applications come at the expense of new challenges to the existing data management schemes:

- Multimedia applications generally require more resources; however, the storage space and processing power are limited in many practical systems; for example, mobile devices and wireless networks (Lim & Hurson, 2002). Due to the large size of multimedia databases and complicated operations of multimedia applications, new methods are needed to facilitate efficient accessing and processing of multimedia data while considering the technological constraints (Bourgeois, Mory & Spies, 2003).
- There is a gap between user perception and physical representation of multimedia data. Users often browse and desire to access multimedia data at the object level ("entities" such as human beings, animals or buildings). However, the existing multimedia-retrieval systems tend to represent multimedia data based on their lower-level features ("characteristics" such as color patterns and textures), with less emphases on combining these features into objects (Hsu, Chua & Pung, 2000). This representation gap often leads to unexpected retrieval

results. The representation of multimedia data according to a human's perspective is one of the focuses in recent research activities; however, no existing systems provide automated identification or classification of objects from general multimedia collections (Kim & Kim, 2002).

- The collections of multimedia data are often diverse and poorly indexed (Huang et al., 2002). In a distributed environment, due to the autonomy and heterogeneity of data sources, multimedia objects are often represented in heterogeneous formats (Kwon, Choi, Bisdikian & Naghshineh, 2003). The difference in data formats further leads to the difficulty of incorporating multimedia objects within a unique indexing framework (Auffret et al., 1999).
- Last but not least, present research on content-based multimedia retrieval is based on features. These features are extracted from the audio/video streams or image pixels, with the empirical or heuristic selection, and then combined into vectors according to the application criteria (Hershey & Movellan, 1999). Due to the application-specific multimedia formats, this paradigm of multimedia data management lacks scalability, accuracy, efficiency and robustness (Westermann & Klas, 2003).

Representation: The Foundation of Multimedia Data Management

Successful storage and access of multimedia data, especially in a distributed heterogeneous database environment, require careful analysis of the following issues:

- Efficient representation of multimedia entities in databases
- Proper indexing architecture for the multimedia databases
- Proper and efficient technique to browse and/or query objects in multimedia database systems.

Among these three issues, multimedia representation provides the foundation for indexing, classification and query processing. The suitable representation of multimedia entities has significant impact on the efficiency of multimedia indexing and retrieval (Huang et al., 2003). For instance, object-level representation usually provides more convenient content-based indexing on multimedia data than pixel-level representation (Kim & Kim, 2002). Similarly, queries are resolved within the representation domains of multimedia data, either at the object level or pixel level (Hsu et al., 2000). The nearest-neighbor searching schemes are usually based on careful analysis of multimedia representation – the knowledge of data contents and organization in multimedia systems (Yu & Zhang, 2000; Li et al., 2003).

The remaining part of this article is organized into three sections: First, we offer the background and related work. Then, we introduce the concepts of semantic-based multimedia representation approach and compare it with the existing non-semantic-based approaches. Finally, we discuss the future trends in multimedia representation and draw this article into a conclusion.

BACKGROUND

Preliminaries of Multimedia Representation

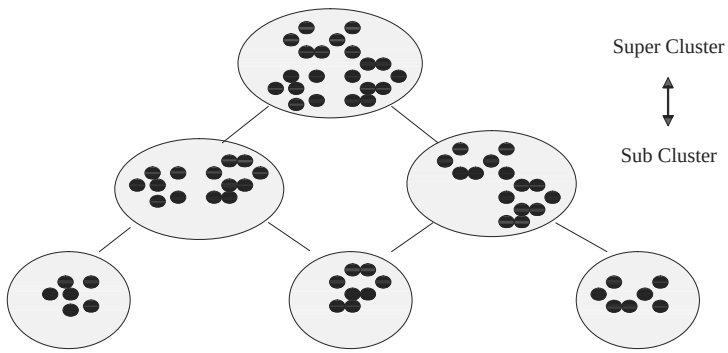
The main goal of multimedia representation is to obtain a concise content description during the analysis of multimedia objects. Representation approaches as advanced in the literature are classified into four groups: clustering-based, representative-region-based, decision-tree-based and annotation-based.

Clustering-Based Approach

The clustering-based approach recursively merges content-similar multimedia objects into clusters with human intervention or automated classification algorithms while obtaining the representation of these multimedia objects. There are two types of clustering schemes: *supervised* and *unsupervised* (Kim & Kim, 2002). The supervised clustering scheme utilizes the user's knowledge and input to cluster multimedia objects, so it is not a general-purpose approach. As expected, the unsupervised clustering scheme does not need interaction with the user. Hence, it is an ideal way to cluster unknown multimedia data automatically (Heisele & Ritter, 1999). Here we only discuss the unsupervised clustering scheme, because of its advantages.

In the clustering-based approach, the cluster of a multimedia object indicates its content (Rezaee, Zwet & Lelieveldt, 2000). The clusters are organized in a hierarchical fashion – a super cluster may be decomposed into several sub clusters and represented as the union of sub clusters (Figure 1). New characteristics are employed in the decomposition process to indicate the differences between sub clusters. Consequently, a sub cluster inherits the characteristics from its super cluster while maintaining its individual contents (Huang et al., 2003).

Figure 1. The decomposition of clusters



Representative-Region-Based Approach

The representative-region-based approach selects several representative regions from a multimedia object and constructs a simple description of this object based on the selected regions. The representative regions are some small areas with the most notable characteristics of the whole object. In case of an image, the representative regions can be areas that the color changes markedly, or areas that the texture varies greatly and so forth.

The representative-region-based approach is performed as a sequence of three steps:

- **Region selection:** The original multimedia object consists of many small regions. Hence, the selection of representative regions is the process of analyzing the changes in those small regions. The difference with the neighboring regions is quantified as a numerical value to represent a region. Finally, based on such a quantitative value, the regions are ordered, and the most notable regions are selected.
- **Function application:** The foundation of the function application process is the Expectation Maximization (EM) algorithm (Ko,& Byun, 2002). The EM algorithm is used to find the maximum likelihood function estimates when the multimedia object is represented by a small number of selected regions. The EM algorithm is divided into

two steps: E-step and M-step. In the E-step, the features for the unselected regions are estimated. In the M-step, the system computes the maximum-likelihood function estimates using the features obtained in the E-step. The two steps alternate until the functions are close enough to the original features in the unselected regions.

- **Content representation:** The content representation is the process that integrates the selected regions into a simple description that represents the content of the multimedia object. It should be noted that the simple description is not necessarily an exhaustive representation of the content. However, as reported in the literature, the overall accuracy of expressing multimedia contents is acceptable (Jing, Li, Zhang & Zhang, 2002).

Decision-Tree-Based Approach

The decision-tree-based approach is the process of obtaining content of multimedia objects through decision rules (MacArthur, Brodley & Shyu, 2000). The decision rules are automatically generated standards that indicate the relationship between multimedia features and content information. In the process of comparing the multimedia objects with decision rules, some tree structures – decision trees – are constructed (Simard, Saatchi & DeGrandi, 2000).

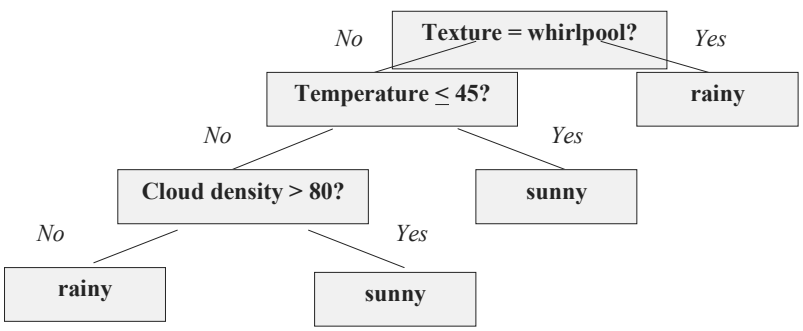
Table 1. The features of cloud images

Temperature	Cloud density	Texture	Weather
45	90	plain	rainy
50	60	whirlpool	rainy
65	75	plain	sunny
38	80	plain	rainy
77	50	plain	sunny
53	85	plain	rainy
67	100	whirlpool	rainy

The decision-tree-based approach is mostly applicable in application domains where decision rules can be used as standard facts to classify the multimedia objects (Park, 1999). For example, in a weather forecasting application, the satellite-cloud images are categorized as rainy and cloudy according to features such as cloud density and texture. Different combinations of feature values are related to different weathers (Table 1). A series of decision rules are derived to indicate these relationships (Figure 2). And the final conclusions are the contents of the multimedia objects.

The decision-tree-based approach can improve its accuracy and precision as the number of analyzed multimedia objects increases (Jeong & Nedevschi, 2003). Since the decision rules are obtained from statistical analysis of multimedia objects, more sample objects will result in improved accuracy (MacArthur et al., 2000).

Figure 2. A decision tree for predicting weathers from cloud images



Annotation-Based Approach

Annotation is the descriptive text attached to multimedia objects. Traditional multimedia database systems employ manual annotations to facilitate content-based retrieval (Benitez, 2002). Due to the explosive expansion of multimedia applications, it is both time-consuming and impractical to obtain accurate manual annotations for every multimedia object (Auffret et al., 1999). Hence, automated multimedia annotation is becoming a hotspot in recent research literature. However, even though humans can easily recognize the contents of multimedia data through browsing, building an automated system that generates annotations is very challenging. In a distributed heterogeneous environment, the heterogeneity of local databases introduces additional complexity to the goal of obtaining accurate annotations (Li et al., 2003).

Semantic analysis can be employed in annotation-based approach to obtain extended content description from multimedia annotations. For instance, an image containing "flowers" and "smiling faces" may be properly annotated as "happiness." In addition, a more complex concept may be deduced from the combination of several simpler annotations. For example, the combination of "boys," "playground" and "soccer" may express the concept "football game."

Comparison of Representation Approaches

The different rationales of these multimedia-representation approaches lead to their strengths and weaknesses in different application domains. Here these approaches are compared under the consideration of various performance merits (Table 2).

The approaches do not consider the semantic contents that may exist in the multimedia objects. Hence, they are collectively called "non-semantic-based" approaches. Due to the lack of semantic analysis, they usually have the following limitations:

- **Ambiguity:** The multimedia contents are represented as numbers that are not easily understood or modified.
- **Lack of robustness and scalability:** Each approach is suitable for some specific application domains, and achieves the best performance only when particular data

Table 2. Comparison of representation approaches

Performance Merit	Clustering	Representative Region	Decision Tree	Annotation
Rationale	Searching pixel-by-pixel, recognizing all details	Selecting representative regions	Treating annotations as multimedia contents	Using annotations as standard facts
Reliability & Accuracy	Reliable and accurate	Lack of robustness	Depending on the accuracy of annotations	Robust and self-learning
Time Complexity	Exhaustive, very time consuming	Most time is spent on region selection	Fast text processing	Time is spent on decision rules and feedback
Space Complexity	Large space requirement	Relatively small space requirement	Very small storage needed	Only need storage for decision rules
Application Domain	Suitable for all application domains	The objects that can be represented by regions	Need annotations as basis	Restricted to certain applications
Implementation Complexity	Easy to classify objects into clusters	Difficult to choose proper regions	Easily obtaining content from annotations	Difficult to obtain proper decision rules

formats are considered. None of them has the capability of accommodating multimedia data of any format from heterogeneous data sources.

MAIN FOCUS OF THE ARTICLE

The limitations of non-semantic-based approaches lead to the research on semantic-based multimedia-representation methods. One of the promising models in the literature is the summary-schemas model (SSM).

Summary Schemas Model

The SSM is a content-aware organization prototype that enables imprecise queries on distributed heterogeneous data sources (Ngamsuriyaroj, 2002). It provides a scalable content-aware indexing method based on the hierarchy of summary schemas, which comprises three major components: a thesaurus, a collection of autonomous local nodes and a set of summary-schemas nodes (Figure 3).

The thesaurus provides an automatic taxonomy that categorizes the standard accessing terms and defines their semantic relationships. A local node is a physical database containing the multimedia data. With the help of the thesaurus, the data items in local databases are classified into proper categories and represented with abstract and semantically equivalent summaries. A summary-schemas node is a virtual entity concisely describing the semantic contents of its child/children node(s). More detailed descriptions can be found in Jiao and Hurson (2004).

To represent the contents of multimedia objects in a computer-friendly structural fashion, the SSM organizes multimedia objects into layers according to their semantic contents. A multimedia object – say, an image – can be considered as the combination of a set of elementary entities, such as animals, vehicles and buildings. And each elementary entity can be described using some logic predicates that indicate the mapping of the elementary entity on different features. For instance, the visual elementary objects in Figure 4 are dog and cat. The possible color is grey, white, blue or brown. The texture pattern is $texture_1$, $texture_2$ or $texture_3$. Hence, the example image in Figure 4 can be represented as the combination of visual objects, colors and textures, such as (cat $\wedge$ brown $\wedge$ t3) $\vee$ (dog $\wedge$ grey $\wedge$ t1).

Figure 3. Summary Schemas Model

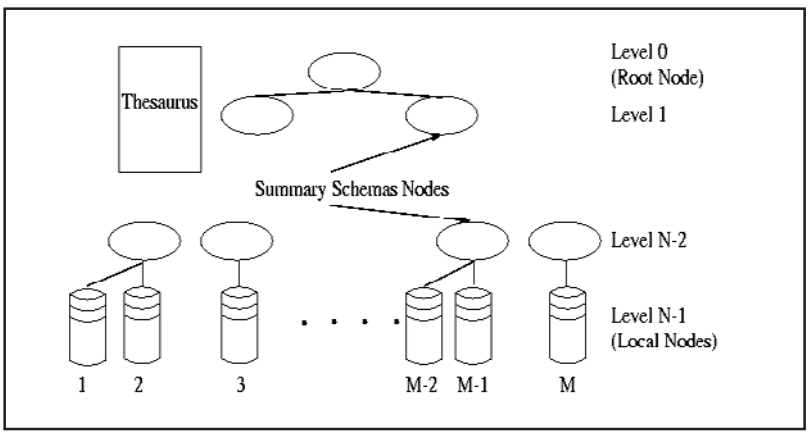
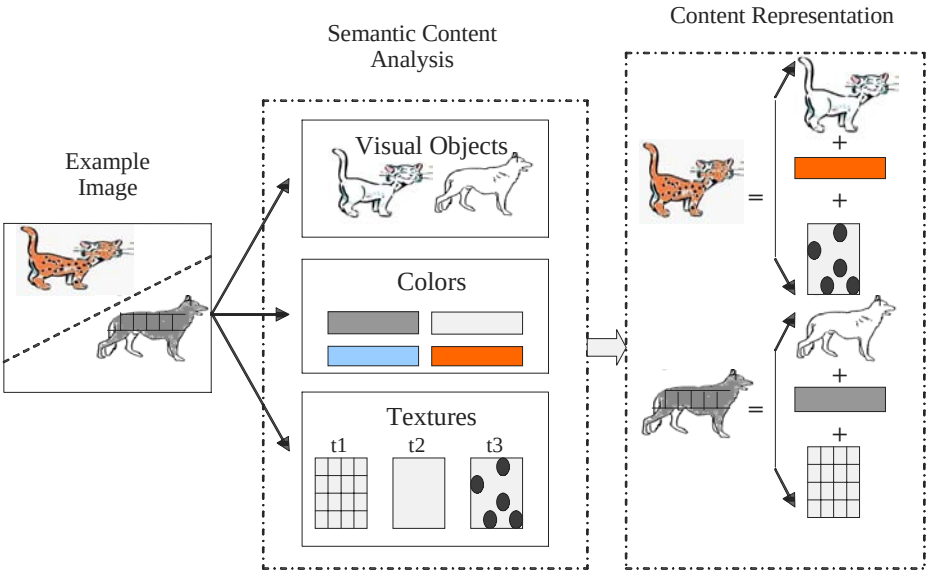


Figure 4. Semantic content components of image objects

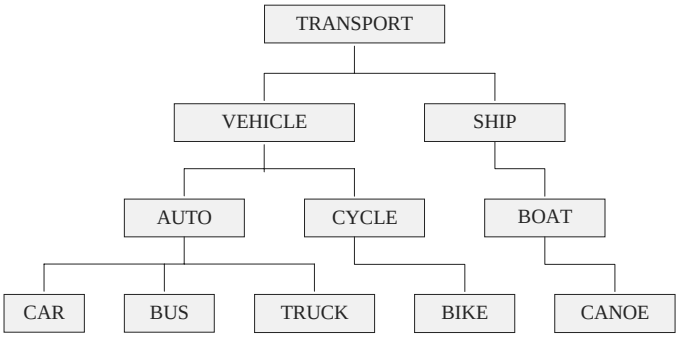


Non-Semantic-Based Methods vs. Semantic-Based Scheme

In contrast with the multimedia-representation approaches mentioned earlier, the SSM employs a unique semantic-based scheme to facilitate multimedia representation and organization. A multimedia object is considered as a combination of logic terms that represents its semantic content. The analysis of multimedia contents is then converted to the evaluation of logic terms and their combinations. This content-representation approach has the following advantages:

- The semantic-based descriptions provide a convenient way of representing multimedia contents precisely and concisely. Easy and consistent representation of the elementary objects based on their semantic features simplifies the content representation of complex objects using logic computations – the logic representation of multimedia contents is often more concise than feature vector, which is widely used in non-semantic-based approaches.
- Compared with non-semantic-based representation, the semantic-based scheme integrates multimedia data of various formats into a unified logical format. This also allows the SSM to organize multimedia objects regardless of their representation (data formats such as MPEG) uniformly, according to their contents. In addition, different media types (video, audio, image and text) can be integrated under the SSM umbrella, regardless of their physical differences.
- The semantic-based logic representation provides a mathematical foundation for operations such as similarity comparison and optimization. Based on the equivalence of logic terms, the semantically similar objects can be easily found and grouped into same clusters to facilitate data retrieval. In addition, mathematical techniques can be used to optimize the semantic-based logic representation of multimedia entities – this by default could result in better performance and space utilization.
- The semantic-based representation scheme allows one to organize multimedia objects

Figure 5. The SSM hierarchy for multimedia objects



in a hierarchical fashion based on the SSM infrastructure (Figure 5). The lowest level of the SSM hierarchy comprises multimedia objects, while the higher levels consist of summary schemas that abstractly describe the semantic contents of multimedia objects. Due to the descriptive capability of summary schemas, this semantic-based method normally achieves more representation accuracy than non-semantic-based approaches.

FUTURE TRENDS AND CONCLUSION

The literature has reported considerable research on multimedia technologies. One of the fundamental research areas is the content representation of multimedia objects. Various non-semantic-based multimedia-representation approaches have been proposed in the literature, such as clustering-based approach, representative-region-based approach, decision-tree-based approach and annotation-based approach. Recent research results also show some burgeoning trends in multimedia-content representation:

- Multimedia-content processing through cross-modal association (Westermann,& Klas, 2003; Li et al., 2003).
- Content representation under the consideration of security (Adelsbach et al., 2003; Lin & Chang, 2001).
- Wireless environment and its impact on multimedia representation (Bourgeois et al., 2003; Kwon et al., 2003).

This article briefly overviewed the concepts of multimedia representation and introduced a novel semantic-based representation scheme – SSM. As multimedia applications keep proliferating through the Internet, the research on content representation will become more and more important.

REFERENCES

Adelsbach, A., Katzenbeisser, S., & Veith, H. (2003). Watermarking schemes provably secure against copy and ambiguity attacks. *ACM Workshop on Digital Rights Management,* 111-119.

Auffret, G., Foote, J., Li, & Shahraray C. (1999). Multimedia access and retrieval (panel session): The state of the art and future directions. *ACM Multimedia, 1,* 443-445.

Benitez, A.B. (2002). Semantic knowledge construction from annotated image collections. *IEEE Conference on Multimedia and Expo, 2,* 205-208.

Bourgeois, J., Mory, E., & Spies, F. (2003). Video transmission adaptation on mobile devices. *Journal of Systems Architecture, 49*(1), 475-484.

Cox, I.J., Miller, M.L., & Minka, T.P. (2000). The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. *IEEE Transactions on Image Processing, 9*(1), 20-37.

Heisele, B., & Ritter, W. (1999). Segmentation of range and intensity image sequences by clustering. *International Conference on Information Intelligence and Systems,* 223-225.

Hershey, J., & Movellan, J. (1999). Using audio-visual synchrony to locate sounds. *Advances in Neural Information Processing Systems,* 813-819.

Hsu, W., Chua, T.S., & Pung, H.K. (2000). Approximating content-based object-level image retrieval. *Multimedia Tools and Applications, 12*(1), 59-79.

Huang, Y., Chang, T., & Huang, C. (2003). A fuzzy feature clustering with relevance feedback approach to content-based image retrieval. *IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems,* 57-62.

Jeong, P., & Nedevschi, S. (2003). Intelligent road detection based on local averaging classifier in real-time environments. *International Conference on Image Analysis and Processing,* 245-249.

Jiao, Y. & Hurson, A.R. (2004). Application of mobile agents in mobile data access systems – A prototype. *Journal of Database Management, 15*(4), 2004.

Jing, F., Li, M., Zhang, H., & Zhang, B. (2002). Region-based relevance feedback in image retrieval. *IEEE Symposium on Circuits and Systems,* 26-29.

Kim, J.B., & Kim, H.J. (2002). Unsupervised moving object segmentation and recognition using clustering and a neural network. *International Conference on Neural Networks, 2,* 1240-1245.

Ko B., & Byun, H. (2002). Integrated region-based retrieval using region's spatial relationships. *International Conference on Pattern Recognition,* 196-199.

Kwon, T., Choi, Y., Bisdikian, C., & Naghshineh, M. (2003). Qos provisioning in wireless/mobile multimedia networks using an adaptive framework. *Wireless Networks,* 51-59.

Li, B., Goh, K., & Chang, E.Y. (2003). Confidence-based dynamic ensemble for image annotation and semantics discovery. *ACM Multimedia,* 195-206.

Li, D., Dimitrova, N., Li, M., & Sethi, I.K. (2003). Multimedia content processing through cross-modal association. *ACM Multimedia,* 604-611.

Lim, J.B., & Hurson, A.R. (2002). Transaction processing in mobile, heterogeneous database systems. *IEEE Transaction on Knowledge and Data Engineering, 14*(6), 1330-1346.

Lin, C., & Chang, S. (2001). SARI: Self-authentication-and-recovery image watermarking system. *ACM Multimedia,* 628-629.

MacArthur, S.D., Brodley, C.E., & Shyu, C. (2000). Relevance feedback decision trees in content-based image retrieval. *IEEE Workshop on Content-based Access of Image and Video Libraries,* 68-72.

Ngamsuriyaroj, S., Hurson, A.R., & Keefe, T.F. (2002). Authorization model for summary schemas model. *International Database Engineering and Applications Symposium,* 182-191.

Park, I.K. (1999). Perceptual grouping of 3D features in aerial image using decision tree classifier. *International Conference on Image Processing, 1,* 31-35.

Rezaee, M.R., Zwet, P.M., & Lelieveldt, B.P. (2000). A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. *IEEE Transactions on Image Processing, 9*(7), 1238-248.

Simard, M., Saatchi, S.S., & DeGrandi, G. (2000). The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *IEEE Transactions on Geoscience and Remote Sensing, 38*(5), 2310–2321.

Westermann, U., & Klas, W. (2003). An analysis of XML database solutions for management of MPEG-7 media descriptions. *ACM Computing Surveys,* 331-373.

Yu, D., & Zhang, A. (2000). Clustertree: Integration of cluster representation and nearest neighbor search for image databases. *IEEE Conference on Multimedia and Expo, 3,* 1713-1716.

KEY TERMS

Annotation: Descriptive text attached to multimedia objects.

Cluster: A group of content-similar multimedia objects.

Decision Rule: Automatically generated standards that indicate the relationship between multimedia features and content information.

Elementary Entity: Data entities that semantically represent basic objects.

Representative Region: Areas with the most notable characteristics of a multimedia object.

Semantic-Based Representation: Describing multimedia content using semantic terms.

Summary-Schemas Model: A content-aware organization prototype that enables imprecise queries on distributed heterogeneous data sources.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 687-695, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.2
Multimedia for Mobile Devices

Kevin Curran
University of Ulster, Ireland

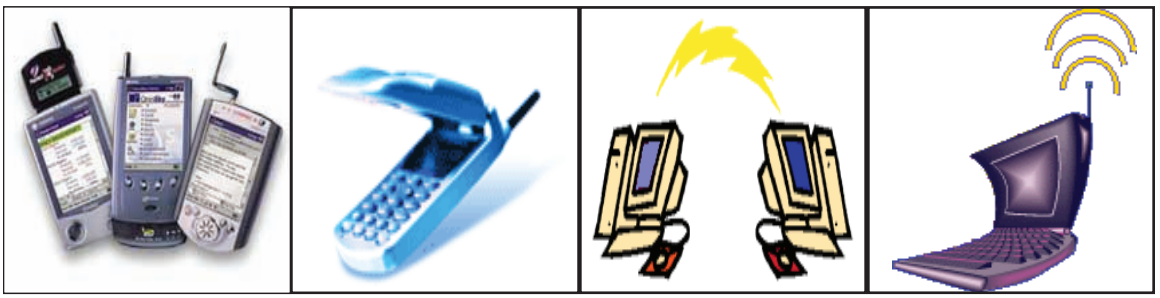
INTRODUCTION

Mobile communications is a continually growing sector in industry and a wide variety of visual services such as video-on-demand have been created, which are limited by low-bandwidth network infrastructures. The distinction between mobile phones and personal device assistants (PDA's) has already become blurred with pervasive computing being the term coined to describe the tendency to integrate computing and communication into everyday life. New technologies for connecting devices like wireless communication and high bandwidth networks make the network connections even more heterogeneous. Additionally, the network topology is no longer static, due to the increasing mobility of users. Ubiquitous computing is a term often associated with this type of networking.

BACKGROUND

The creation of low bit rate standards such as H.263 (Harrysson, 2002) allows reasonable quality video through the existing Internet and is an important step in paving the way forward. As these new media services become available, the demand for multimedia through mobile devices will invariably increase. Corporations such as Intel do not plan to be left behind. Intel has created a new breed of mobile chip code named Banias. Intel's president and chief operating officer Paul Otellino states that "eventually every single chip that Intel produces will contain a radio transmitter that handles wireless protocols, which will allow users to move seamlessly among networks. Among our employees this initiative is affectionately referred to as 'radio free Intel.'" Products such as Real Audio and IPCast for streaming media are also becoming increasingly common, however, multimedia, due to its timely nature requires guarantees different in nature with regards to delivery of data from TCP traffic such as HTTP requests. In addition, multimedia applications increase the set of requirements in terms of throughput, end-to-end delay, delay jitter, and clock synchronisation. These requirements may not all be directly met by the networks therefore end-system protocols enrich network services

Figure 1. (a) PDAs; (b) mobiles; (c) desktops; (d) laptops



(a) (b) (c) (d)

to provide the quality of service (QoS) required by applications. It is argued here that traditional monolithic protocols are unable to support the wide range of application requirements on top of current networks (ranging from 9600 baud modems up to gigabit networks) without adding overhead in the form of redundant functionality for numerous combinations of application requirements and network infrastructures. In ubiquitous computing, software is used by roaming users interacting with the electronic world through a collection of devices ranging from handhelds such as PDAs (Figure 1a) and mobile phones (Figure 1b) to personal computers (Figure 1c) and laptops (Figure 1d).

The Java language, thanks to its portability and support for code mobility, is seen as the best candidate for such settings (Kochnev & Terekhov, 2003; Román et al., 2002). The heterogeneity added by modern smart devices is also characterised by an additional property, which is that many of these devices are typically tailored to distinct purposes. Therefore, not only memory and storage capabilities differ widely, but local device capabilities, in addition to the availability of resources changing over time (e.g., a global positioning satellite (GPS) system cannot work indoors unless one uses specialised repeaters (Jee, Boo, Choi, & Kim, 2003). Therefore a need exists for middleware to be aware of these pervasive computing properties. With regards to multimedia, applications that use group communication (e.g., video conferencing) mechanisms must be able to scale from small groups with few members, up to groups with thousands of receivers (Tojo, Enokido, & Takizawa, 2003).

The protocols underlying the Internet were not designed for the latest cellular type networks with their low bandwidth, high error losses, and roaming users, thus many "fixes" have arisen to solve the problem of efficient data delivery to mobile resource constrained devices (Saber & Mirenkov, 2003). Mobility requires adaptability meaning that systems must be location-aware and situation-aware taking advantage of this information in order to dynamically reconfigure in a distributed fashion (Matthur & Mundur, 2003; Solon, McKevitt, & Curran, 2003). However, situations, in which a user moves an end-device and uses information services can be challenging. In these situations, the placement of different cooperating parts is a research challenge.

ENABLING TECHNOLOGIES FOR MOBILE MULTIMEDIA

In 1946, the first car-based telephone was set up in St. Louis, MO, USA. The system used a single radio transmitter on top of a tall building. A single channel was used, and therefore a button was pushed to talk and released to listen (Tanenbaum, 2005). This half duplex system is still used by modern day CB-radio systems used by police and taxi operators. In the 60's, the system was improved to a two-channel system called improved mobile

telephone system (IMTS). The system could not support many users as frequencies were limited. The problem was solved by the idea of using cells to facilitate the re-use of frequencies. More users can be supported in such a cellular radio system. It was implemented for the first time in the advanced mobile phone system (AMPS). Wide-area wireless data services have been more of a promise than a reality. It can be argued that success for wireless data depends on the development of a digital communications architecture that integrates and interoperates across regional-area, wide-area, metropolitan-area, campus-area, in-building, and in-room wireless networks.

The convergence of two technological developments has made mobile computing a reality. In the last few years, the UK and other developed countries have spent large amounts of money to install and deploy wireless communication facilities. Originally aimed at telephone services (which still account for the majority of usage), the same infrastructure is increasingly used to transfer data. The second development is the continuing reduction in the size of computer hardware, leading to portable computation devices such as laptops, palmtops, or functionally enhanced cell phones. Unlike second-generation cellular networks, future cellular systems will cover an area with a variety of non-homogeneous cells that may overlap. This allows the network operators to tune the system layout to subscriber density and subscribed services. Cells of different sizes will offer widely varying bandwidths: very high bandwidths with low error rates in pico-cells, very low bandwidths with higher error rates in macro-cells as illustrated in Table 1. Again, depending on the current location, the sets of available services might also differ.

Unlike traditional computer systems characterised by short-lived connections that are bursty in nature, Streaming audio/video sessions are typically long lived (the length of a presentation) and require continuous transfer of data. Streaming services will require, by today's standards, the delivery of enormous volumes of data to customer homes. For examples, entertainment NTSC video compressed using the MPEG standards requires bandwidths between 1.5 and 6 Mb/s. Many signalling schemes have been developed that can deliver data at this rate to homes over existing communications links (Forouzan, 2002). Some signalling schemes suitable for high-speed video delivery are:

- **ADSL:** The asymmetrical digital subscriber loop (ADSL) (Bingham, 2000) takes advantage of the advances in coding to provide a customer with a downstream wideband signal, an upstream control. The cost to the end-user is quite low in this scheme, as it requires little change to the existing equipment.

Table 1. Characteristics of various wireless networks

Type of Network	Bandwidth Latency	Latency	Mobility	Typical Video Performance	Typical Audio Performance
In-Room/Building (Radio Frequency Infrared)	>> 1 Mbps RF: 2-20 Mbps IR: 1-50 Mbps	<< 10 ms	Pedestrian	2-Way, Interactive, Full Frame Rate (Compressed)	High Quality, 16 bit samples, 22 KHz rate
Campus-Area Packet Relay	Approx. 64 kbps	Approx. 100 ms	Pedestrian	Medium Quality Slow Scan	Medium Quality Reduced Rate
Wide-Area (Cellular, PCS)	19.2 kbps	> 100 ms	Pedestrian/ Vehicular	Video Phone or Freeze Frame	Asynchronous "Voice Mail"
Regional-Area (LEO/VSAT DBS)	Asymmetric Up/Dn 100 bps to 4.8 kbps 12 Mbps	>> 100 ms	Pedestrian/ Vehicular Stationary	Async Video Playback	Asynchronous "Voice Mail"

- **CATV:** Cable TV (CATV) (Forouzan, 2002) uses a broadband coaxial cable system and can support multiple MPEG compressed video streams. CATV has enormous bandwidth capability and can support hundreds of simultaneous connections. Furthermore, as cable is quite widely deployed, the cost of supporting Video-on-demand and other services is significantly lower. However, it requires adaptation to allow bi-directional signalling in the support of interactive services.

A cellular wireless network consists of fixed based stations connecting mobile devices through a wired backbone network where each mobile device establishes contact through their local base stations. The available bandwidth on a wireless link is limited and channels are more prone to errors. It is argued that future evolution of network services will be driven by the ability of network elements to provide enhanced multimedia services to any client anywhere (Harrysson, 2002). Future network elements must be capable of transparently accommodating and adjusting to client and content heterogeneity. There are benefits to filtering IP packets in the wireless network so that minimal application data is carried to the mobile hosts to preserve radio resources and prevent the overloading of mobile hosts with unnecessary information and ultimately wasteful processing. A proxy is an intermediary component between a source and a sink, which transforms the data in some manner. In the case of mobile hosts, a proxy is often an application that executes in the wired network to support the host. This location is frequently the base station, the machine in the wired network that provides the radio interface. As the user moves, the proxy may also move to remain on the communication path from the mobile device to the fixed network. The proxy hides the mobile from the server, which thinks that it communicates with a standard client (i.e., a PC directly connected to the wired network) (Kammann & Blachnitzky, 2002).

Wireless links are characterised by relatively low bandwidth and high transmission error rates (Chakravorty & Pratt, 2002). Furthermore, mobile devices often have computational constraints that preclude the use of standard Internet video formats on them thus by placing a mobile transcoding proxy at the base station (BS), the incoming video stream can be transcoded to a lower bandwidth stream, perhaps to a format more suitable to the nature of the device, and control the rate of output transmission over the wireless link (Joshi, 2000).

Figure 2 illustrates a scenario where a transcoding gateway is configured to transcode MPEG streams to H.261. In the architecture, the transcoding gateway may also simply forward MPEG or H.261 packets to an alternate session (in both directions) without performing transcoding. Figure 3 illustrates locations in which intelligence about available network services may be placed. Client may utilise this network knowledge to select the most appropriate server and mechanism in order to obtain appropriate content. As an alternative, this knowledge (and the associated burden) could be entirely or partially transferred to the individual servers or could reside inside the network.

Image transcoding is where an image is converted from one format to another (Vetro, Sun,

Figure 2. A transcoding proxy

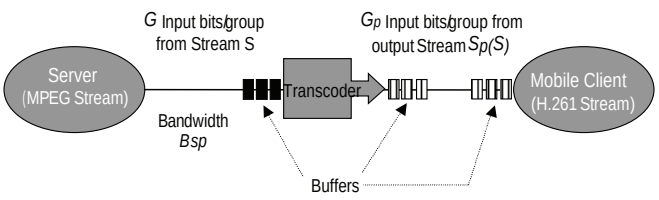
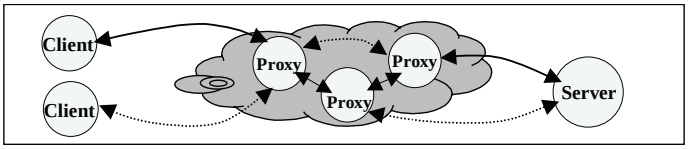


Figure 3. Variations in client-server connectivity



& Wang, 2001). This may be performed by altering the Qscale (basically applying compression to reduce quality). This is sometimes known as simply resolution reduction. Another method is to scale down the dimensions of the image (spatial transcoding) (Chandra, Gehani, Schlatter Ellis, & Vahdat, 2001) so reduce the overall byte size (e.g., scaling a 160Kb frame by 50% to 32KB). Another method known as temporal transcoding is where frames are simply dropped (this can sometimes be known as simply rate reduction). While another method may be simply to transcode the image to greyscale, which may be useful for monochrome PDA,'s (again this transcoding process results in reduced byte size of the image or video frame). Recently there has been increased research into intelligent intermediaries). Support for streaming media in the form of media filters has also been proposed for programmable heterogeneous networking. Canfora, Di Santo, Venturi, Zimeo, and Zito (2005) propose multiple proxy caches serving as intelligent intermediaries, improving content delivery performance by caching content. A key feature of these proxies is that they can be moved and re-configured to exploit geographic locality and content access patterns thus reducing network server load. Proxies may also perform content translation on static multimedia in addition to distillation functions in order to support content and client heterogeneity (Yu, Katz, & Laksham, 2005). Another example is fast forward networks broadcast overlay architecture where there are media bridges in the network, which can be used in combination with RealAudio or other multimedia streams to provide an application layer multicast overlay network. One could adopt the view at this time that "boxes" are being placed in the network to aid applications.

MOBILE IP

Mobile IP is an extension to IP, which allows the transparent routing of IP datagrams to mobile nodes. In mobile IP, each host has a home agent (i.e., a host connected to the sub network the host is attached to). The home agent holds responsibility for tracking the current point of attachment of the mobile device so when the device changes the network it is connected to it has to register a new care-of address with the home agent. The care-of address can be the address of a foreign agent (e.g., a wireless base station node) that has agreed to provide services for the mobile or the new IP address of the mobile (if one is dynamically assigned by its new network). Traffic to mobiles is always delivered via home agents, and then tunnelled to the care-of address. In the case of a foreign agent care-of address, traffic is forwarded to the mobile via the foreign agent. Traffic from the mobile does not need to travel via the home agent but can be sent directly to (previously) correspondents. Correspondent hosts do not need to be mobile IP enabled or even have to know the location of the mobile if a home agent acts as an intermediary, thus forwarding of packets to the current address of the mobile is transparent for other hosts [Dixit02]. The home agent redirects packets from the home network to the care-of address by creating a new IP header, which contains the mobile's care-of address as the destination IP address. This new header encapsulates the

original packet, causing the mobile node's home address to have no effect on the encapsulated packet's routing until it reaches the care-of address. When the mobile leaves the service area of its current foreign agent and registers with a new foreign agent, the home agent must be informed about the change of address. In the process of handoff, the mobile may lose connectivity for a short period of time.

FUTURE TRENDS

Mobile phone technologies have evolved in several major phases denoted by "generations" or "G" for short. Three generations of mobile phones have evolved so far, each successive generation more reliable and flexible than the previous. The first of these is referred to as the first generation or 1G. This generation was developed during the 1980s and early 1990s and only provided an analog voice service with no data services available (Bates, 2002). The second generation or 2G of mobile technologies used circuit-based digital networks. Since 2G networks are digital they are capable of carrying data transmissions, with an average speed o f around 9.6K bps (bits per second). Because 2G networks can support the transfer of data, they are able to support Java enabled phones. Some manufacturers are providing Java 2 Micro Edition (J2ME) (Knudsen & Li, 2005) phones for 2G networks though the majority are designing their Java enabled phones for the 2.5G and 3G networks, where the increased bandwidth and data transmission speed will make these applications more usable (Hoffman, 2002). These are packet based and allow for "always on" connectivity. The third generation of mobile communications (3G) is digital mobile multimedia offering broadband mobile communications with voice, video, graphics, audio and other forms of information. 3G builds upon the knowledge and experience derived from the preceding generations of mobile communication, namely 2G and 2.5G although 3G networks use different transmission frequencies from these previous generations and therefore require a different infrastructure (Camarillo& Garcia-Martin, 2005). These networks will improve data transmission speed up to 144K bps in a high-speed moving environment, 384K bps in a low-speed moving environment, and 2Mbps in a stationary environment. 3G services see the logical convergence of two of the biggest technology trends of recent times, the Internet and mobile telephony. Some of the services that will be enabled by the broadband bandwidth of the 3G networks include:

- Downloadable and streaming audio and video
- Voice over Internet protocol (VoIP)
- Send and receive high quality colour images
- Electronic agents are self-contained programs that roam communications networks delivering /receiving messages or looking for information or services.
- Downloadable software—Potential to be more convenient than conventional methods of distributing software as the product arrives in minutes
- Capability to determine geographic position of a mobile device using the global positioning system (GPS) (Barnes et al., 2003)

3G will also facilitate many other new services that have not previously been available over mobile networks due to the limitations in data transmission speeds. These new wireless applications will provide solutions to companies with distributed workforces, where employees need access to a wide range of information and services via their corporate intranets, when they are working offsite with no access to a desktop (Camarillo & Camarillo, 2005).

4G technology stands to be the future standard of wireless devices. The Japanese company NTT DoCoMo is testing 4G communication at

100 Mbit/s while moving, and 1 Gbit/s while stationary. NTT DoCoMo plans on releasing the first commercial network in 2010. Despite the fact that current wireless devices seldom utilize full 3G capabilities, there is a basic attitude that if you provide the pipeline then services for it will follow.

CONCLUSION

Flexible and adaptive frameworks are necessary in order to develop distributed multimedia applications in such heterogeneous end-systems and network environments. The processing capability differs substantially for many of these devices with PDA's being severely resource constrained in comparison to leading desktop computers. The networks connecting these devices and machines range from GSM, Ethernet LAN, and Ethernet 802.11 to Gigabit Ethernet. Networking has been examined at a low-level micro-protocol level and again from a high-level middleware framework viewpoint. Transcoding proxies were introduced as a promising way to achieving dynamic configuration, especially because of the resulting openness, which enables the programmer to customize the structure of the system and other issues regarding mobility were also discussed.

REFERENCES

Barnes, J. Rizos, C., Wang, J., Small, D., Voigt, G., & Gambale, N. (2003). LocataNet: A new positioning technology for high precision indoor and outdoor positioning. In *Proceedings of (Institute of Navigation) ION GPS/GNSS 2003,* Oregon Convention Center, Portland, Oregon, September 9-12, 2003

Bates, J. (2002). *Optimizing voice transmission in ATM/IP mobile networks.* McGraw-Hill Telecom Engineering, London, UK.

Bingham, J. (2000). *ADSL, VDSL, and multicarrier modulation* (1st ed.). Wiley-Interscience, Manchester, UK.

Camarillo, G., & Garcia-Martin, M. (2005). *The 3G IP multimedia subsystem (IMS): Merging the Internet and the cellular worlds (2nd ed.).* John Wiley and Sons Ltd, London, UK.

Canfora, G., Di Santo, G., Venturi, G., Zimeo, E., & Zito, M. (2005). Migrating Web application sessions in mobile computing. In *Proceedings of the International World Wide Web Conference*, 2005

Chandra,, S., Gehani, A., Schlatter Ellis, C., & Vahdat, A. (2001). Transcoding characteristics of web images. In *Proceedings of the SPIE Multimedia Computing and Networking Conference*, January 2001.

Chakravorty, R., & Pratt, I. (2002). WWW Performance over GPRS. The *4th IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002),* Stockholm, Sweden, September 9-11, 2002

Dixit, S., & Prasad, R. (2002). *Wireless IP and building the mobile Internet*. Artech House Universal Personal Communications Series, Norwood, MA, USA.

Feng, Y., & Zhu, J. (2001). *Wireless Java programming with J2ME* (1st ed.). Sams Publishing.

Forouzan, B. (2002). *Data communications and networking* (2nd ed.). McGraw-Hill Publishers, London, UK.

Harrysson, A. (2002). Industry challenges for mobile services. The *4th IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002),* Stockholm, Sweden, September 9-11, 2002

Hoffman, J. (2002). *GPRS demystified* (1st e.). McGraw-Hill Professional, London, UK.

Jee, G., Boo, S., Choi, J., & Kim, H. (2003). An indoor positioning using GPS repeater. In *Proceedings of (Institute of Navigation) ION GPS/GNSS 2003,* Oregon Convention Center, Portland, Oregon, September 9-12, 2003

Joshi, A. (2000). On Proxy Agents, Mobility and Web Access. *Mobile Networks and Applications, 5*(4), 233-241

Kammann, J., & Blachnitzky, T. (2002). Split-proxy concept for application layer handover in mobile communication systems. The *4th IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002),* Stockholm, Sweden, September 9-11, 2002

Knudsen, J., & Li, S. (2005). *Beginning J2ME: From novice to professional*. APress, New York, NY, USA.

Kochnev, D., & Terekhov, A. (2003). Surviving Java for mobiles. *IEEE Pervasive Computing, 2*(2), 90-95.

Matthur, A., & Mundur, P. (2003). Congestion adaptive streaming: An integrated approach. DMS'2003—The *9th International Conference on Distributed Multimedia Systems,* Florida International University Miami, Florida, USA, September 24-26, 2003

Román, M., Hess, C., Cerqueira, R., Ranganathan, A., Campbell, R., & Nahrstedt. K. (2002). A middleware infrastructure for active spaces. *IEEE Pervasive Computing, 1*(4), 74-83.

Saber, M., & Mirenkov. N. (2003). A multimedia programming environment for cellular automata systems. DMS'2003—The *9th International Conference on Distributed Multimedia Systems,* Florida International University Miami, Florida, USA, September 24-26, 2003

Solon, T., McKevitt, P., & Curran, K. (2003). Telemorph—Bandwidth determined mobile multimodal presentation. *IT&T 2003 – Information Technology and Telecommunications,* Letterkenny Institute of Technology, Co. Donegal, Ireland. 22-23rd October, 2003

Tanenbaum, A. (2005). *Computer networks* (5th ed.). Prentice Hall, New Jersey, USA.

Tojo, T., Enokido, T., & Takizawa, M. (2003). Notification-based QoS control protocol for group communication. DMS'2003—The *9th International Conference on Distributed Multimedia Systems,* Florida International University Miami, Florida, USA, September 24-26, 2003

Vetro, A., Sun, H., & Wang, Y. (2001). Object-based transcoding for adaptable video content delivery. *IEEE Trans. Circuits and System for Video Tech*nology, *11*(2), 387-401, March 2001.

Yu, F., Katz, R., & Laksham, T. (2005). Efficient multi-match packet classification and lookup with TCAM. *IEEE Micro magazine, 25*(1), 50-59, February 2005.

KEY TERMS

Bandwidth: The amount of data that can be transferred from one point to another, usually between a Web server and a Web browser; It is a measure of the range of frequencies a transmitted signal occupies. In digital systems, bandwidth is the data speed in bits per second. In analog systems, bandwidth is measured in terms of the difference between the highest-frequency signal component and the lowest-frequency signal component.

Broadband: The telecommunication that provides multiple channels of data over a single communications medium.

Cellular Network: A cellular wireless network consists of fixed based stations connecting mobile devices through a wired backbone network where each mobile device establishes contact through their local base stations. The available bandwidth on a wireless link is limited and channels are more prone to errors.

Content: Data that an encoder or server streams to a client or clients. Content can originate from live audio or live video presentation, stored audio or video files, still images, or slide shows. The content must be translated from its original state into a Windows Media format before a Windows Media server can stream it. Windows Media servers can stream live streams or stored Window Media files as content.

Encoding: Encoding accomplishes two main objectives: (1) it reduces the size of video and audio files, by means of compression, making Internet delivery feasible, and (2) it saves files in a format that can be read and played back on the desktops of the targeted audience. Encoding may be handled by a software application or by specialised hardware with encoding software built in.

Media: A term with many different meanings, in the context of *streaming media*, it refers to video, animation, and audio. The term "media" may also refer to something used for storage or transmission, such as tapes, diskettes, CD-ROMs, DVDs, or networks such as the Internet.

Multiple Bit Rate Video: The support of multiple encoded video streams within one media stream. By using multiple bit rate video in an encoder, you can create media-based content that has a variety of video streams at variable bandwidths ranging from, for example, 28.8 Kbps through 300 Kbps, as well as a separate audio stream. After receiving this multiple encoded stream, the server determines which bandwidth to stream based on the network bandwidth available. Multiple bit rate video is not supported on generic HTTP servers.

Streaming Video: A sequence of moving images that are transmitted in compressed form over the Internet and displayed by a viewer as they arrive; is usually sent from pre-recorded video files, but can be distributed as part of a live broadcast feed.

Third Generation Mobile Communications (3G): The third generation of mobile communications (3G) is digital mobile multimedia offering broadband mobile communications with voice, video, graphics, audio, and other forms of information.

Web Casting: The technique of broadcasting media over an intranet, extranet, or the Internet.

Web Server Streaming: Another term for HTTP streaming, pseudo-streaming, or progressive download

ENDNOTES

1 www.realaudio.com
2 www.ipcast.com
3 http://www.3gnewsroom.com

This work was previously published in Encyclopedia of Internet Technologies and Applications, edited by M. Freire and M. Pereira, pp. 323-330, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.3
Multimedia Contents for Mobile Entertainment

Hong Yan
City University of Hong Kong, Hong Kong
University of Sydney, Australia

Lara Wang
Tongji University, China

Yang Ye
Tongji University, China

INTRODUCTION

Electronic mobile devices are becoming more and more powerful in terms of memory size, computational speed, and color display quality. These devices can now perform many multimedia functions, including rendering text, sound, images, color graphics, video, and animation. They can provide users with great entertainment values, which were only possible with more expensive and bulky equipment before.

Technological advances in computer and telecommunications networks have also made mobile devices more useful for information exchange among users. As a result, a number of new mobile products and services, such as multimedia messages services (MMSs) and online games, can be offered to users by the industry.

It is commonly believed that "contents are the king" for multimedia products and services. As the mobile handsets and networks become more and more advanced, there is a stronger and stronger demand for high-quality multimedia contents to be used in new hardware systems. Content creation involves both computer technology and artistic creativity. Due to a large number of users, mobile entertainment has become an important part of the so-called creative industry that is booming in many countries.

In this article, we provide an overview of multimedia contents for mobile entertainment applications. The objective is for the readers to become familiar with basic multimedia technology, and the concepts and commonly used terms. Our focus will be on multimedia signal representation, processing, and standards.

SOUND

The original sound, such as speech from humans, can be represented as a continuous function $s(t)$. The sound can be recorded using electronic devices and stored on magnetic tapes. It can also be transmitted through telecommunications systems. The traditional telephone is used to send and receive waveform information of sound signals. The function $s(t)$ here is called an analog audio signal. This signal can be sampled every T seconds, or $f_s = 1/T$ times per second. The output signal is called a discrete-time signal. Each data sample in a discrete-time signal can have an arbitrary value. The sample values can be quantized so that we can represent them using a limited number of bits in the computer. These two processes, sampling and quantization, are called digitization, which can be achieved using an analog-to-digital converter (ADC) (Chapman & Chapman, 2004; Gonzalez & Woods, 2002; Mandal, 2003; Ohm, 2004). A signal $s(n)$ that is discrete both in time and in amplitude is called a digital signal. To render a digital sound signal, we must use a digital-to-analog converter (DAC) and send the analog signal to an electronic speaker.

The parameter $f_s = 1/T$ is called the sampling frequency. For telephone applications, usually f_s = 8000Hz, and for audio CD usually f_s = 44100Hz. To achieve stereo effects, the audio CD has two channels of data. There is only one channel in telephone applications. The sampling frequency f_s must be greater or equal to twice the signal bandwidth in order to reconstruct or recover the analog signal correctly. This is called the Nyquist sampling criterion. In telephone systems, a sound signal may have to be filtered to remove high-frequency components so that the sampling criterion is satisfied. This is why audio CDs have a higher sound quality than telephones.

The sound information can be stored as raw digital data. Some sound files, such as WAVE files (with extension ".wav") used on PCs, store raw sound data. This kind of format requires large storage space but has the advantage, since there is no information loss and the data can be accessed easily and quickly. To reduce the amount of data for storage and transmission, sound data are often compressed. Commonly used sound data compression methods include the m-law transformation, adaptive differential pulse code modulation (ADPCM), and Moving Picture Experts Group (MPEG) audio compression (Chapman & Chapman, 2004; Mandal, 2003; MPEG, n.d.; Ohm 2004).

Currently, MPEG Audio Layer 3 (MP3) is a very popular sound data compression technique for mobile devices. MP3 is also a well-known sound file format. During data compression in MP3, a sound signal is decomposed into 32 frequency bands, and psychoacoustic models are used to determine the masking level and bit allocation pattern for each band. Modified discrete cosine transform (MDCT) is used to compress the data. The discrete cosine transform (DCT) has the so-called energy compact property—that is, it is able to pack most of the energy of a signal in a small number of DCT coefficients. For example, if $s(0) = 2$, $s(1) = 5$, $s(2) = 7$, and $s(3) = 6$, then the DCT coefficients are $S(0) = 10$, $S(1) = 3.15$, $S(2) = 2$, and $S(3) = 0.22$ (Mandal, 2003). In this case, from $S(0)$ to $S(3)$, the coefficients become smaller and smaller. We can simply retain the low-frequency components $S(0)$ and $S(1)$ and discard high-frequency ones $S(2)$ and $S(3)$ to obtain an approximation of the original signal. In practice, we can allocate more bits to code $S(0)$ and less and less bits for $S(1)$ to $S(3)$ to achieve a high compression ratio and at the same time maintain good signal quality.

Short sound files can be completely downloaded to mobile devices before being played. To reduce waiting time for downloading long sound files, audio streaming technology can be used (Austerberry 2005). In a streaming system, audio data is transmitted through a network and played by a mobile device as the data become available. That is, a sound does not have to be completely

stored in the mobile device before being played. Steaming is useful if a large amount of data need to be received by a mobile device or live broadcasting is required.

We have focused on how to process sound waveform information above. In fact, sound can also be generated according to its parameters or a set of instructions. The musical instrument digital interface (MIDI) standard is used for such purpose (Chapman & Chapman, 2004; Mandal, 2003). A MIDI file contains information on what kind of instruments, such as different types of pianos, should be used and how they should be played. Several instruments can be arranged in different channels and played at the same time. MIDI files are much smaller than waveform-based sound files for music and is widely used for ring tones on mobile phones.

IMAGES

The imaging ability of mobile devices has been improved rapidly in recent years. Now most new mobile phones are equipped with digital cameras and can take pictures with millions of pixels. Software programs are available to edit an image, such as to enhance its contrast and change its color appearance. Mobile devices can also be used to send or receive images and browse the Web.

An important parameter for images is the resolution, which is closely related but should not be confused with the image size. Image resolution is usually measured by dots per inch (dpi). Higher resolution for the same physical area of an object would generate a larger image than lower resolution, but a large image does not necessarily mean a high resolution as it depends on the area that the image covers. The concept of resolution is often used in image printing and scanning. Typically, laser printers have a resolution of 300dpi or 600dpi and fax documents have resolutions from 100dpi to 300dpi.

A digital image can be considered as a two-dimensional (2D) discrete function $i(x, y)$. Like sound data, images need to be compressed to save storage space and transmission time. There are a number of methods that can be used to compress an image. Most methods are designed to reduce the spatial redundancy so that an image can be represented using a smaller amount of data. The most popular technique used for image compression is the Joint Photographic Experts Group (JPEG) standard (Gonzalez & Woods 2002). In JPEG-based compression, an image is divided into small, square blocks, and each block is transformed using the two-dimensional DCT (2D-DCT). Similar to audio data compression, the energy of a smooth image is concentrated in low-frequency components of the 2D-DCT coefficients. By allocating more bits to a small number of low-frequency components than to a large number of high-frequency components, we can achieve effective data compression.

In the JPEG2000 standard, the 2D discrete wavelet transform (2D-DWT) is used for image compression. In this method, an image is decomposed into several sub-bands, and different quantization schemes are used for different sub-bands. The 2D-DWT usually performs better—that is, it can provide a higher quality for similar compression ratio or a higher compression ratio for similar image quality than the 2D-DCT.

The JPEG and JPEG200 standards are usually used to provide lossy compression with a high compression ratio, although they can also be used for lossless compression. In lossy compression, the decompressed image is only an approximation of the original one, and as a result the image may appear to be blocky and blurred. The quality of an image from lossy compression can be improved using a number of techniques (Liew & Yan, 2004; Weerasingher, Liew, & Yan, 2002; Zou & Yan, 2005). In lossless compression, we can reconstruct or recover the original image exactly. Commonly used lossless compression methods include graphic interchange format (GIF), tagged

image file format (TIFF), and portable network graphics (PNG). A GIF image can only show 256 colors and is especially useful for logos, buttons, borders, and simple animation on Web pages. GIF is a patented technology. In TIFF, image information is associated with different tags, which can be defined by users. TIFF is widely used for scanned office documents. PNG, which can support true colors, has been developed to be a royalty-free alternative to GIF. Similar to GIF, PNG can provide background transparency, but PNG does not support animation.

GRAPHICS

An image can also be generated from a set of drawing operations, similar to the way music is generated from MIDI instructions. Images represented by pixel values are called bitmaps, while those obtained from drawings are called vector graphics. The bitmap format is better for natural objects, such as human faces and outside sceneries, whereas the vector format is better for computer-generated objects, such as line drawings and industrial designs.

On computers and mobile devices, different operating or window systems provide different graphics utilities as software development tools. In general, they should all have graphics functions for drawing line shapes, such as lines, polygons, and circles, and displaying bitmaps. Some systems provide more advanced functions, such as geometric shape transformations, color gradients, and spline curve and surface displays. Currently, there are many proprietary mobile operating systems developed by different manufacturers. This makes it difficult to port graphics applications from one phone model to another model. However, a few systems have been adopted by more and more manufacturers. They include Symbian, Microsoft Windows Mobile OS, and Linux-based mobile OS.

There are also several graphics formats that are supported by many mobile operating systems. They include proprietary formats Java 2 Micro Edition (J2ME) and Flash, and royalty-free and open standard scalable vector graphics (SVG). J2ME is a general purpose programming language for small electronic devices and contains many useful graphics functions (Edward, 2003; Wells, 2004). It has gained widespread use in mobile phones, especially for games.

Flash was developed by Macromedia, which is now part of Adobe Systems. In addition to graphics, Flash can also be used to present other types of multimedia data, such as text, sound, and animation. The output file of Flash is called an SWF movie, which contains definition tags and control tags. The definition tags describe the objects in a movie, such as text, shapes, bitmaps, sound, and sprites. The control tags define how and when an object should be transformed and displayed. SWF movie files use a very compact binary format, so they can be transmitted over the Internet quickly. They have found many applications to Web page design and are now used more and more in mobile phones. Macromedia Flash provides an authoring tool to design SWF movies. One can also create a SWF movie using a computer program according to the SWF file format.

SVG is an open graphics standard developed by the World Wide Web Consortium (W3C). It is royalty free and vendor independent. SVG uses text format and provides a language to define graphics display, in a way similar to the HyperText Markup Language (HTML) that defines text display. SVG can be used to describe vector shapes, text, bitmap images, and animation. Since vector graphics can be easily scaled larger or smaller, SVG files can be used for different resolutions, such as in printing which requires a high resolution, and in displays on mobile devices which have low resolution. A number of graphics software packages, such as CorelDraw and Adobe Illustrator, can output SVG files.

VIDEO

A digital video contains a sequence of images $v(x, y, n)$, where x and y are spatial coordinates and n represents time. For each n value or a time sample, we have a frame of video or an image. In video compression, we can explore both spatial or intra-frame redundancy and temporal or inter-frame redundancy. Spatially, a frame can be compressed just like an image. Temporally, a frame is similar to its proceeding frame, so we can expect a higher compression ratio for video than for each image separately and independently.

A straightforward way of reducing the temporal redundancy is to use the proceeding frame and approximation of the current frame. A better approximation can be achieved if we take into account movements of the objects in the image sequence. We can divide an image into small blocks and search in the proceeding frame for a closest shifted version of each block. Then, we only need to code the difference between the block under consideration and its closest match in the previous frame, and the shift between the two blocks. This procedure is called motion estimation.

Extensive research has been carried out in the field of video processing on how to estimate and compensate motions efficiently.

A series of MPEG standards have been developed for video compression (MPEG, n.d.; Watkinson, 2004). MPEG-1 was developed for coding of moving pictures and associated audio for digital storage media, such as video CD and MP3, at up to 1.5 Mbit/s. It uses block motion compensation and the DCT to code the residual image. MPEG-2 provides generic coding of moving pictures and associated audio information for bit rates from 1.5 to 80 Mbit/s. It is an extension of MPEG-1, allows combinations of video and audio streams, and supports different packet formats for data transmission. MPEG-2 products include digital TV setup boxes and DVDs. MPEG-4 provides standardized technology for video and audio data storage, transmission, and content access and manipulation on digital TV, interactive graphics applications, and the World Wide Web. In addition to video and audio coding, it supports creation of synthetic objects. For example, mesh models can be used for human face animation. MPEG-7 is a standard for describing multimedia content, including images, graphics, 3D models, audio, speech, video, and the way various data should be combined in a multimedia presentation. MPEG-21, which is still under development, defines the multimedia framework for different users, including content creators, producers, distributors, and service providers, to access, exchange, manipulate, and trade a large variety of multimedia items.

The International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) has developed a series of standards H.26x for video phone and videoconference applications. H.261 supports data rates as multiples of 64Kbit/s, that is, $p\times$64Kbit/s, where $1 \le p \le 30$. H.261 only supports two image frame sizes, common interchange format (CIF) or 352×288 pixels, and quarter common interchange format (QCIF) or 176×144 pixels. H.262 is the same as the video part of MPEG-2. H.263 provides a number of improvements over H.261, MPEG-1, and MPEG-2. For example, it uses better methods for motion compensation, offers higher video quality, and supports more image sizes. H.264 is the same as MPEG-4 Part 10, advanced video coding (AVC), and is jointly developed by the ITU-T Video Coding Experts Group (VCEG) and MPEG. H.264/MPEG-4 AVC employs a number of new techniques, such as variable block-size motion compensation (VBSMC), to improve the compression performance. It is also more flexible under a variety of network environments. H.263 and H.264 have been used in 3GP movies for second-generation (2G) and third-generation (3G) mobile phones.

Similar to audio, video can also be streamed (Austerberry, 2005). This is especially useful for video phone calling, videoconferencing, and live

broadcasting. Current 3G mobile networks already offer video calls and many video-based entertainment programs, such as news and sports, which were only possible through TV before.

ANIMATION

Animation means presentation of a sequence of artificially created images. In video, the images are obtained from a camera, while in animation, the images are drawn by hand or generated by the computer. Moving pictures in animation are often synchronized with audio to create movies.

Cartoon movies had been used for entertainment long before the digital computer was invented. A cartoon film must contain 24 picture frames every second to show smooth motions (Chapman & Chapman, 2004). This means 86,400 pictures for a one-hour movie; obviously it is very labor intensive. The labor cost can be reduced using cel animation and key-frame animation techniques. In cel animation, the still background is drawn only once and the moving part is drawn frame by frame, each on a cel, a sheet of transparent material, which is placed on the background picture. In key-frame animation, a motion is decomposed into important key frames and in-betweens. Experienced animators are assigned to draw the key frames and junior animators the in-betweens. Now using the computer, the in-betweens can often be generated using pattern matching and interpolation techniques.

Simple animation can be displayed on the computer or mobile devices by simply going though a sequence of images. GIF provides such function. It is useful for small animated pictures, which does not require many colors or audio. It is widely used for Web page design and for MMS on mobile phones.

Simple animation can also be generated using morphing techniques. In Macromedia Flash, one shape can be morphed to another shape specified by a set of parameters. Each shape is described by its edges and color information. In this method, key frames are drawn according to shape specifications, and in-betweens are automatically generated by the computer through the morphing process.

Three-dimensional (3D) animation has been used in many digital entertainment products and services. In 3D animation, an object is often represented using a mesh model. It is shaded according to light and camera settings, and may be superimposed with a texture. Now there are a number of 3D design packages available for PCs as well as powerful workstations. Virtual reality (VR) systems make extensive use of 3D graphics and animation. The Virtual Reality Modeling Language (VRML) was developed to support VR on the Web (Ames, Nadeau, & Moreland, 1997).

An interesting and challenging task in computer animation is to automatically animate the human and animal faces (Huang & Yan, 2002, 2003; Parke & Waters, 1996; Yan, 2001). This involves synchronizing the mouth movement with voice and generating different facial expressions. We have recently developed a real-time lip synchronization and facial animation system, which has already been used by several Internet and communications companies for 3G and MMS applications on mobile phones (Tang, Liew, & Yan 2005). Our system can create many virtual characters with different styles. It can match voice signals with lip/mouth shapes smoothly and automatically. The lip-sync can be done in real time (e.g., 20 frames per second) for several languages, including Chinese (Cantonese and Mandarin), English, French, German, Japanese, and Spanish. In addition, a virtual character can express his/her emotion with different kinds of facial expressions. Examples of the movies produced by our system can be found on http://www.HyperAcademy.com or simply http://www.hy8.com.

TECHNOLOGY TRENDS

Mobile technology is rapidly advancing. We expect to see many new or enhanced mobile entertainment products and services in the next several years. First of all, future mobile devices will have higher computational power for its processors, higher resolution for digital cameras, and higher data rates through the telecommunications network. Currently, there is criticism for 3G mobile phones that the battery does not last long for displaying movies. To support multimedia functions for a longer period of usage, mobile devices need improved technologies for reduced power assumption and better batteries.

Traditionally, most multimedia contents based on graphics, video, and animation are produced for TV and other large display screens. Simple image down sampling of the images can cause visibility problems. So future multimedia editing and authoring tools should take into account the legibility of the material for mobile devices. An interesting research topic in image processing and pattern recognition is how to reduce the size of a bitmap image optimally so that the output is most legible. This requires sophisticated algorithms for feature detection and pattern extraction from the image.

More and more multimedia contents will become available for mobile devices in the future. Mobile gaming is one of the rapidly growing areas. There will also be increased use of mobile phones for music, movies, advertisements, news reports, storytelling, finance, weather and traffic information, and educational and training programs, and for access to the Web, most of which are similar to the capabilities and functions provided by TV and desktop computers. For Web applications, better user interface and page layout will be needed so that the users can find information needed easily. We also expect increased applications of video phone calls, videoconferencing, and video e-mail messages.

A more challenging task for a mobile device to perform is to recognize its user's voice, speech, face, and even facial expressions and emotions, and interact with the user. Considerable progress has been made in the past in these areas, but the recognition reliability still needs significant improvement. These problems can be solved to some extent under restricted conditions. For mobile phones, it is difficult to control the background noise or appearance for normal usages, so more robust pattern recognition techniques must be developed to solve these problems.

ACKNOWLEDGMENTS

This work is supported by a research grant from City University of Hong Kong (Project 9610034).

REFERENCES

Ames, A. L., Nadeau, D. R., & Moreland, J. L. (1997). *VRML 2.0 sourcebook*. New York: John Wiley & Sons.

Austerberry, D. (2005). *The technology of video and audio streaming.* Burlington, MA: Focal Press, Elsevier.

Chapman, N., & Chapman, J. (2004). *Digital multimedia*. Chichester, UK: John Wiley & Sons.

Edward, J. (2003). *J2ME: The complete reference*. New York: McGraw-Hill.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing.* Upper Saddle River, NJ: Prentice Hall.

Huang, D., & Yan, H. (2002). Modeling and animation of human expressions using NURBS curves based on facial anatomy. *Signal Processing: Image Communications, 17,* 457-465.

Huang, D., & Yan, H. (2003). NURBS curve controlled modelling for facial animation. *Computers & Graphics, 27,* 373-385.

Liew, A., & Yan, H. (2004). Blocking artifacts suppression in block-coded images using overcomplete wavelet representation. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(4), 450-461.

Mandal, M. K. (2003). *Multimedia signals and systems.* Boston: Kluwer Academic.

MPEG. (n.d.). *Homepage.* Retrieved from http://www.chiariglione.org/mpeg/

Ohm, J.-R. (2004). *Multimedia communication technology, representation, transmission and identification of multimedia signals.* New York: Springer.

Parke, F. I., & Waters, K. (1996). *Computer facial animation.* Wellesley, MA: A.K. Peters.

Tang, J. S. S., Liew, A., & Yan, H. (2005). Human face animation based on video analysis, with applications to mobile entertainment. *Journal of Mobile Multimedia, 1*(2), 132-147.

Watkinson, J. (2004). *The MPEG handbook: MPEG-1, MPEG-2, MPEG-4.* Burlington, MA: Focal Press, Elsevier.

Weerasinghe, C., Liew, A., & Yan, H. (2002). Artifact reduction in compressed images based on region homogeneity constraints using projection on to convex sets algorithm. *IEEE Transactions on Circuits and Systems for Video Technology, 12*(10), 891-897.

Wells, M. J. (2004). *J2ME game programming.* Boston: Thomson Course Technology.

Yan, H. (2001). Image analysis for digital media applications. *IEEE Computer Graphics and Applications, 21*(1), 18-26.

Zou, J. J., & Yan, H. (2005). A delocking method for BDCT compressed images based on adaptive projections. *IEEE Transactions on Circuits and Systems for Video Technology, 15*(3), 430-435.

KEY TERMS

Audio Signal Processing: Acquisition, compression, enhancement, filtering, transformation, and transmission of sound data.

Computer Graphics: Modeling and rendering of two-dimensional or three-dimensional objects on the computer.

Digital Entertainment: Providing multimedia materials, such as music and movies, on digital devices, such as computers and mobile phones.

Facial Animation: Modeling and displaying talking human faces with different expressions.

Image Processing: Acquisition, compression, enhancement, filtering, transformation, and transmission of pictures.

Lip-Synchronization: Matching the mouth shape created by the computer with the voice signal.

Multimedia Content: Digital data of text, music, graphics, images, and video

Video Processing: Acquisition, compression, ehancement, filtering, transformation, and transmission of movies.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 669-674, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.4
Multimedia Information Design for Mobile Devices

Mohamed Ally
Athabasca University, Canada

INTRODUCTION

There is a rapid increase in the use of mobile devices such as cell phones, tablet PCs, personal digital assistants, Web pads, and palmtop computers by the younger generation and individuals in business, education, industry, and society. As a result, there will be more access of information and learning materials from anywhere and at anytime using these mobile devices. The trend in society today is learning and working on the go and from anywhere rather than having to be at a specific location to learn and work. Also, there is a trend toward ubiquitous computing, where computing devices are invisible to the users because of wireless connectivity of mobile devices. The challenge for designers is how to develop multimedia materials for access and display on mobile devices and how to develop user interaction strategies on these devices. Also, designers of multimedia materials for mobile devices must use strategies to reduce the user mental workload when using the devices in order to leave enough mental capacity to maximize deep processing of the information. According to O'Malley et al. (2003), effective methods for presenting information on these mobile devices and the pedagogy of mobile learning have yet to be developed. Recent projects have started research on how to design and use mobile devices in the schools and in society. For example, the MOBILearn project is looking at pedagogical models and guidelines for mobile devices to improve access of information by individuals (MOBILearn, 2004). This paper will present psychological theories for designing multimedia materials for mobile devices and will discuss guidelines for designing information for mobile devices. The paper then will conclude with emerging trends in the use of mobile devices.

BENEFITS AND LIMITATIONS OF MOBILE DEVICES

There are many benefits of using mobile devices in the workplace, education, and society. In mobile learning (m-learning), users can access information and learning materials from anywhere and at anytime. There are many definitions of m-learning in the field. M-learning is the use of electronic learning materials with built-in learning strategies for delivery on mobile computing devices to allow access from anywhere and at anytime (Ally, 2004a). Another definition of m-learning

is any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of the learning opportunities offered by mobile technologies (O'Malley et al., 2003). With the use of wireless technology, mobile devices do not have to be physically connected to networks in order to access information. Mobile devices are small enough to be portable, which allows users to take the device to any location to access information or learning materials. Because of the wireless connectivity of mobile devices, users can interact with other users from anywhere and at anytime to share information and expertise, complete a task, or work collaboratively on a project. Mobile devices have many benefits, because they allow for mobility while learning and working; however, there are some limitations of mobile devices that designers must be aware of when designing multimedia materials for delivery on mobile devices.

Some of the limitations of mobile devices in delivering multimedia materials include the small screen size for output of information, small input devices, low bandwidth, and challenges when navigating through the information (Ahonen et al., 2003). Designers of information and learning materials have to be aware of the limited screen size and input device when designing for usability. For example, rather than scrolling for more information on the screen, users of mobile devices must be able to go directly to the information and move back and forth with ease. Information should be targeted to the users' needs when they need it and should be presented efficiently to maximize the display on the mobile device. To compensate for the small screen size of mobile devices, multimedia materials must use rich media to convey the message to the user. For example, rather than present information in textual format, graphics and pictures can be used in such a way to convey the message using the least amount of text. For complex graphics, a general outline of the graphic should be presented on one screen with navigation tools to allow the user to see the details of the graphic on other screens. To present procedures and real-life situations, video clips can be used to present real-life simulations to the user. Also, the interface must be appropriate for individual users and the software system should be able to customize the interface based on individual users' characteristics. When developing multimedia materials for mobile devices, designers must be aware of psychological theories in order to guide the design.

PSYCHOLOGICAL THEORY FOR DEVELOPING MULTIMEDIA MATERIALS FOR MOBILE DEVICES

According to cognitive psychology, learning is an internal process, and the amount learned depends on the processing capacity of the user, the amount of effort expended during the learning process, the quality of the processing, and the user's existing knowledge structure (Ausubel, 1974). These have implications for how multimedia materials should be designed for mobile devices. Designers must include strategies that allow the user to activate existing cognitive structure and conduct quality processing of the information. Mayer et al. (2003) found that when a pedagogical agent was present on the screen as instruction was narrated to students, students who were able to ask questions and receive feedback interactively perform better on a problem-solving transfer test when compared to students who only received on-screen text with no narration. It appears that narration by a pedagogical agent encouraged deep processing, which resulted in higher-level learning. According to Paivio's theory of dual coding, memory is enhanced when information is represented both in verbal and visual forms (Paivio, 1986). Presenting materials in both textual and visual forms will involve more processing, resulting in

better storage and integration in memory (Mayer et al., 2004). Tabbers et al. (2004) found that in a Web-based multimedia lesson, students who received visual cues to pictures scored higher on a retention test when compared to students who did not receive the cues for the pictures. Also, strategies can be included to get the user to retrieve existing knowledge to process the information presented. For example, a comparative advance organizer can be used to activate existing knowledge structure to process the incoming information, or an expository advance organizer can be presented and stored in memory to help incorporate the details in the information (Ally, 2004a; Ausubel, 1974).

Constructivism is a theory of learning that postulates that learners are active during the learning process, and that they use their existing knowledge to process and personalize the incoming information. Constructivists claim that learners interpret information and the world according to their personal realities, and that they learn by observation, processing, and interpretation and then personalize the information into their existing knowledge bases (Cooper, 1993). Users learn best when they can contextualize what they learn for immediate application and to acquire personal meaning. According to Sharples (2000), mobile learning devices allow learners to learn wherever they are located and in their personal context so that the learning is meaningful. Also, mobile devices facilitate personalized learning, since learning is contextualized where learning and collaboration can occur from anywhere and anytime. According to constructivism, learners are not passive during the learning process. As a result, interaction on mobile devices must include strategies to actively process and internalize the information. For example, on a remote job site, a user can access the information using a mobile device for just-in-time training and then apply the information right away. As a result, designers must use instructional strategies to allow users to apply what they learn.

DESIGN GUIDELINES FOR MULTIMEDIA MATERIALS FOR MOBILE DEVICES

Cater for the User of Mobile Devices

- **Design for the User:** One of the variables that designers tend to ignore when they develop multimedia materials for mobile devices is the user of the devices. Different users have different learning styles; some users may be visual, while others may be verbal (Mayer & Massa, 2003). Users have different learning styles and preferences; strategies must be included and information presented in different ways in order to cater to the different learning styles and preferences (Ally & Fahy, 2002). Graphic overviews can be used to cater to users who prefer to get the big picture before they go into the details of the information. For active learners, information can be presented on the mobile device, and then the user can be given the opportunity to apply the information. For the creative users, there must be opportunities to apply the information in real-life applications so that they go beyond what was presented. The multimedia materials and information have to be designed with the user in mind in order to facilitate access, learning, and comprehension. Also, the user should have control of what he or she wants to access in order to go through the multimedia materials based on preferred learning styles and preferences. For users in remote locations with low bandwidth or limited wireless access, information that takes a long time to download should be redesigned to facilitate efficient download.
- **Adapt the Interface to the User:** An interface is required to coordinate interaction between the user and the information. To compensate for the small screen size of the display of the mobile device, the interface of

the mobile device must be designed properly. The interface can be graphical and should present limited information on the screen to prevent information overload in short-term memory. The system should contain intelligent software agents to determine what the user did in the past and to adapt the interface for future interaction with the information. The software system must be proactive by anticipating what the user will do next and must provide the most appropriate interface for the interaction to enhance learning. Users must be able to jump to related information without too much effort. The interface must allow the user to access the information with minimal effort and move back to previous information with ease. For sessions that are information-intense, the system must adjust the interface to prevent information overload. Some ways to prevent information overload include presenting less concepts on one screen or organizing the information in the form of concept maps to give the overall structure of the information and then presenting the details by linking to other screens with the details. The interface also must use good navigational strategies to allow users to move back and forth between displays. Navigation can also be automatic based on the intelligence gathered on the user's current progress and needs.

- **Design for Minimum Input:** Because of the small size of the input device, multimedia materials must be designed to require minimum input from users. Input can use pointing or voice input devices to minimize typing and writing. Because mobile devices allow access of information from anywhere at anytime, the device must have input and output options to prevent distractions when using the mobile devices. For example, if someone is using a mobile device in a remote location, it may be difficult to type on a keyboard or use a pointing device. The mobile technology must allow the user to input data using voice input or touch screen.
- **Build Intelligent Software Agents to Interact with the User:** Intelligent software systems can be built to develop an initial profile of the user and then present materials that will benefit the specific user, based on the user profile. As the intelligent agent interacts with the user, it learns about the user and adapts the format of the information, the interface, and the navigation pattern according to the user's style and needs. Knowing the user's needs and style will allow the intelligent software system to access additional materials from the Internet and other networks to meet the needs of user (Cook et al., 2004).
- **Use a Personalized Conversational Style:** Multimedia information and learning materials can be presented to the user in a personalized style or a formal style. In a learning situation, information should be presented in a personalized style, since the user of the mobile device may be in a remote location and will find this style more connected and personal. Mayer et al. (2004) found that students who received a personalized version of a narrated animation performed significantly better on a transfer test when compared to students who received a non-personalized, formal version of the narrated animation. They claimed that the results from the study are consistent with the cognitive theory of multimedia learning, where personalization results in students processing the information in an active way, resulting in higher-level learning and transfer to other situations.

Design to Improve the Quality of Information Processing on Mobile Devices

- **Chunk Information for Efficient Processing:** Designers of materials for mobile devices must use presentation strategies to enable users to process the materials efficiently because of the limited display capacity of mobile devices and the limited processing capacity of human working memory. Information should be organized or chunked in segments of appropriate and meaningful size to facilitate processing in working memory. An information session on a mobile device can be seen as consisting of a number of information objects sequenced in a predetermined way or sequenced based on the user needs. Information and learning materials for mobile devices should take the form of information and learning objects that are in an electronic format, reusable, and stored in a repository for access anytime and from anywhere (McGreal, 2004). Information objects and learning objects allow for instant assembly of learning materials by users and intelligent software agents to facilitate just-in-time learning and information access. The information can be designed in the form of information objects for different learning styles and characteristics of users (Ally, 2004b). The objects then are tested and placed in an electronic repository for just-in-time access from anywhere and at anytime using mobile devices.
- **Use High-Level Concept Maps to Show Relationships:** A concept map or a network diagram can be used to show the important concepts in the information and the relationship between the concepts rather than present information in a textual format. High-level concept maps and networks can be used to represent information spatially so that students can see the main ideas and their relationships (Novak, Gowin, & Johanse, 1983). Tusack (2004) suggests the use of site maps as the starting point of interaction that users can link back in order to continue with the information or learning session. Eveland et al. (2004) compared linear Web site designs and non-linear Web site designs and reported that linear Web site designs encourage factual learning, while non-linear Web site designs increase knowledge structure density. One can conclude that the non-linear Web site designs show the interconnection of the information on the Web site, resulting in higher-level learning.

EMERGING TRENDS IN DESIGNING MULTIMEDIA MATERIALS FOR MOBILE DEVICES

The use of mobile devices with wireless technology allow access of information and multimedia materials from anywhere and anytime and will dramatically alter the way we work and conduct business and how we interact with each other (Gorlenko & Merrick, 2003). For example, mobile devices can make use of Global Positioning Systems to determine where users are located and connect them with users in the same location so that they can work collaboratively on projects and learning materials. There will be exponential growth in the use of mobile devices to access information and learning materials, since the cost of the devices will be lower than desktop computers, and users can access information from anywhere and at anytime. Also, the use of wireless mobile devices would be more economical, since it does not require the building of the infrastructure to wire buildings. The challenge for designers of multimedia materials for mobile devices is how to standardize the design for use by different types of devices. Intelligent software agents should be

built into mobile devices so that most of the work is done behind the scenes, minimizing input from users and the amount of information presented on the display of the mobile devices. Because mobile devices provide the capability to access information from anywhere, future multimedia materials must be designed for international users.

CONCLUSION

In the past, the development of multimedia materials and mobile devices concentrated on the technology rather than the user. Future development of multimedia materials for mobile devices should concentrate on the user to drive the development and delivery (Gorlenko & Merrick, 2003). Mobile devices can be used to deliver information and learning materials to users, but the materials must be designed properly in order to compensate for the small screen of the devices and the limited processing and storage capacity of a user's working memory. Learning materials need to use multimedia strategies that are information-rich rather than mostly text.

REFERENCES

Ahonen, M., Joyce, B., Leino, M., & Turunen, H. (2003). Mobile learning: A different viewpoint. In H. Kynaslahti, & P. Seppala (Eds.), *Mobile learning* (pp. 29-39). Finland: Edita Publishing Inc.

Ally, M. (2004a). Using learning theories to design instruction for mobile learning devices. *Proceedings of the Mobile Learning 2004 International Conference*, Rome.

Ally, M. (2004b). Designing effective learning objects for distance education. In R. McGreal (Ed.), *Online education using learning objects* (pp. 87-97). London: RoutledgeFalmer.

Ally, M., & Fahy, P. (2002). Using students' learning styles to provide support in distance education. *Proceedings of the Eighteenth Annual Conference on Distance Teaching and Learning*, Madison, Wisconsin.

Ausubel, D.P. (1974). *Educational psychology: A cognitive view.* New York: Holt, Rinehart and Winston.

Cook, D.J., Huber, M., Yerraballi, R., & Holder, L.B. (2004). Enhancing computer science education with a wireless intelligent simulation environment. *Journal of Computing in Higher Education*, *16*(1), 106-127.

Cooper, P.A. (1993). Paradigm shifts in designing instruction: From behaviorism to cognitivism to constructivism. *Educational Technology, 33*(5), 12-19.

Eveland, W.P., Cortese, J., Park, H., & Dunwoody, S. (2004). How website organization influences free recall, factual knowledge, and knowledge structure density. *Human Communication Research, 30*(2), 208-233.

Gorlenko, L., & Merrick, R. (2003). No wires attached: Usability challenges in the connected mobile world. *IBM Systems Journal*, *42*(4), 639-651.

Mayer, R.E., Dow, T.D., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds. *Journal of Educational Psychology, 95*(4), 806-813.

Mayer, R.E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*(2), 389-395.

Mayer, R.E., & Massa, L.J. (2003). Three facets of visual and verbal learners: Cognitive ability,

cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833-846.

McGreal, R. (2004). *Online education using learning objects.* London: Routledge/Falmer.

MOBILearn Leaflet. (2004): Next-generation paradigms and interfaces for technology supported learning in a mobile environment exploring the potential of ambient intelligence. Retrieved September 8, 2004, from *http://www.mobilearn.org/results/results.htm*

Novak, J.D., Gowin, D.B., & Johanse, G.T. (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education, 67,* 625-645.

O'Malley, C., et al. (2003). Guidelines for learning/teaching/tutoring in a mobile environment. Retrieved September 8, 2004, from *http://www.mobilearn.org/results/results.htm*

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers and Education, 34,* 177-193.

Tabbers, H.K., Martens, R.L., & van Merrienboer, J.J.G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74,* 71-81.

Tusack, K. (2004). Designing Web pages for handheld devices. *Proceedings of the 20th Annual Conference on Distance Teaching and Learning,* Madison, Wisconsin.

KEY TERMS

Advance Organizer: A general statement at the beginning of the information or lesson to activate existing cognitive structure or to provide the appropriate cognitive structure to learn the details in the information or the lesson.

Concept Map: A graphic outline that shows the main concepts in the information and the relationship between the concepts.

Intelligent Software Agent: A computer application software that is proactive and capable of flexible autonomous action in order to meet its design objectives set out by the designer. The software learns about the user and adapts the interface and the information to the user's needs and style.

Interface: The components of the computer program that allow the user to interact with the information.

Learning Object: A digital resource that is stored in a repository that can be used and reused to achieve a specific learning outcome or multiple outcomes (Ally, 2004b).

Learning Style: A person's preferred way to learn and process information, interact with others, and complete practical tasks.

Mobile Device: A device that can be used to access information and learning materials from anywhere and at anytime. The device consists of an input mechanism, processing capability, storage medium, and display mechanism.

Mobile Learning (M-Learning): Electronic learning materials with built-in learning strategies for delivery on mobile computing devices to allow access from anywhere and at anytime.

Multimedia: A combination of two or more media to present information to users.

Short-Term Memory: The place where information is processed before the information is transferred to long-term memory. The duration of short-term memory is very short, so information must be processed efficiently to maximize transfer to long-term memory.

Ubiquitous Computing: Computing technology that is invisible to the user because of wireless connectivity and transparent user interface.

User: An individual who interacts with a computer system to complete a task, learn specific knowledge or skills, or access information.

Wearable Computing Devices: Devices that are attached to the human body so that the hands are free to complete other tasks.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 704-709, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.5
Multimedia over Wireless Mobile Data Networks

Surendra Kumar Sivagurunathan
University of Oklahoma, USA

Mohammed Atiquzzaman
University of Oklahoma, USA

ABSTRACT

With the proliferation of wireless data networks, there is an increasing interest in carrying multimedia over wireless networks using portable devices such as laptops and personal digital assistants. Mobility gives rise to the need for handoff schemes between wireless access points. In this chapter, we demonstrate the effectiveness of transport layer handoff schemes for multimedia transmission, and compare with Mobile IP, the network layer-based industry standard handoff scheme.

INTRODUCTION

Mobile computers such as personal digital assistants (PDA) and laptop computers with multiple network interfaces are becoming very common. Many of the applications that run on a mobile computer involve multimedia, such as video conferencing, audio conferencing, watching live movies, sports, and so forth. This chapter deals with multimedia communication in mobile wireless devices, and, in particular, concentrates on the effect of mobility on streaming multimedia in wireless networks.

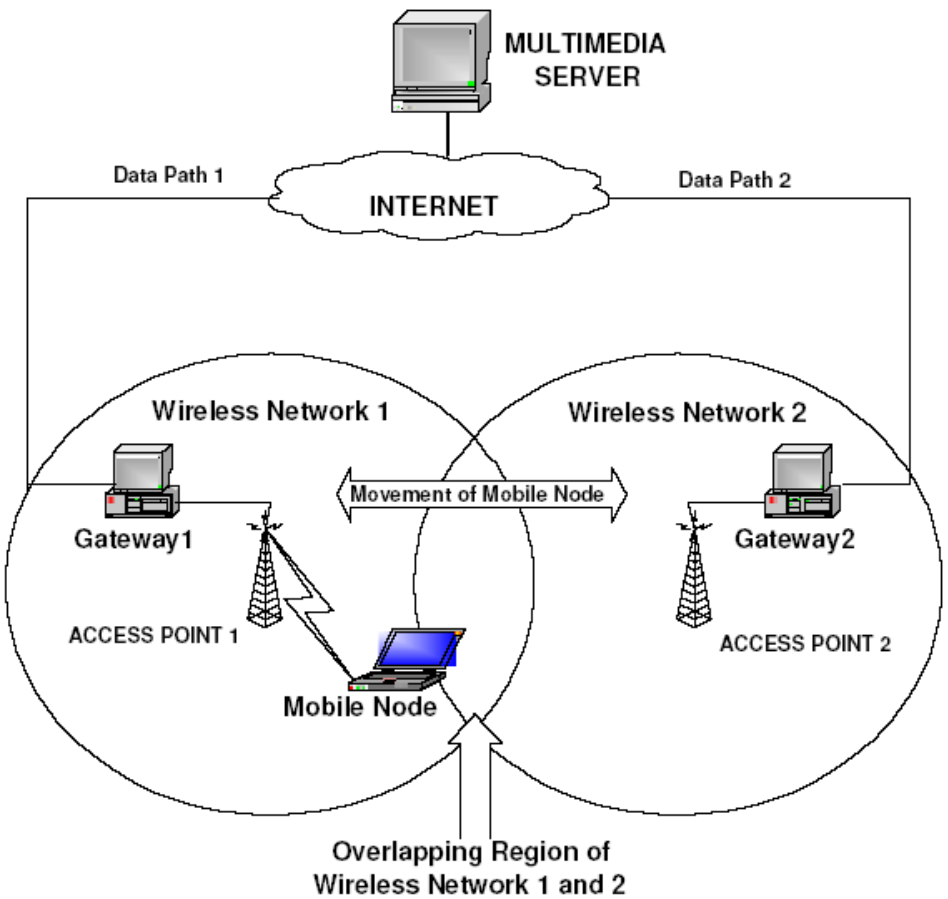
Streaming multimedia over wireless networks is a challenging task. Extensive research has been carried out to ensure a smooth and uninterrupted multimedia transmission to a mobile host (MH) over wireless media. The current research thrust is to ensure an uninterrupted multimedia transmission when the MH moves between networks or subnets. Ensuring uninterrupted multimedia transmission during handoff is challenging because the MH is already receiving multimedia from the network to which it is connected; when it moves into another network, it needs to break the connection with the old network and establish a connection with the new network. Figure 1 shows an MH connected to Wireless Network

1; when it moves, it has to make a connection with the new network, say Wireless Network 2. The re-establishment of a new connection takes a considerable amount of time, resulting in the possibility of interruption and resulting loss of multimedia.

The current TCP/IP network infrastructure was not designed for mobility. It does not support handoff between IP networks. For example, a device running a real-time application, such as video conference, cannot play smoothly when the user hands off from one wireless IP network to another, resulting in unsatisfactory performance to the user.

Mobile IP (MIP) (Perkins, 1996), from the Internet Engineering Task Force (IETF), addresses the mobility problem. MIP extends the existing IP protocol to support host mobility, including handoff, by introducing two network entities: home agent (HA) and foreign agent (FA). The HA and FA work together to achieve host mobility. The correspondent node (CN) always communicates with the mobile node (MN) via its home network address, even though MH may not dwell in the home network. For CN to have seamless access to MN, the MH has to be able to handoff in a timely manner between networks.

Figure 1. Illustration of handoff with mobile node connected to Wireless Network 1



Handoff latency is one of the most important indicators of handoff performance. Large handoff latency degrades performance of real-time applications. For example, large handoff latency will introduce interruption in a video conference due to breaks in both audio and video data transmission. In addition to high handoff latency, MIP suffers from a number of other problems including triangle routing, high signaling traffic with the HA, and so forth. A number of approaches to reduce the MIP handoff latency are given next.

Mobile IP uses only one IP; a certain amount of latency in data transmission appears to be unavoidable when the MH performs a handoff. This is because of MN's inability to communicate with the CN through either the old path (because it has changed its wireless link to a new wireless network) or the new path (because HA has not yet granted its registration request). Thus, MH cannot send or receive data to or from the CN while the MH is performing registration, resulting in interruption of data communication during this time interval. This interruption is unacceptable in a real-world scenario, and may hinder the widespread deployment of real-time multimedia applications on wireless mobile networks. Seamless IP-diversity based generalized mobility architecture (SIGMA) overcomes the issue of discontinuity by exploiting multi-homing (Stewart, 2005) to keep the old data path alive until the new data path is ready to take over the data transfer, thus achieving lower latency and lower loss during handoff between adjacent subnets than Mobile IP.

The *objective* of this chapter is to demonstrate the effectiveness of SIGMA in reducing handoff latency, packet loss, and so forth, for multimedia transmission, and compare with that achieved by Mobile IP. The *contribution* of this chapter is to describe the implementation of a real-time streaming server and client in SIGMA to achieve seamless multimedia streaming during handoff. SIGMA *differs* from previous work in the sense that all previous attempts modified the hardware, infrastructure of the network, server, or client to achieve seamless multimedia transmission during handoff.

The rest of this chapter is organized as follows. Previous work on multimedia over wireless networks is described in the next section. The architecture of SIGMA is described in the third section, followed by the testbed on which video transmission has been tested for both MIP and SIGMA in the fourth section. Results of video over MIP and SIGMA and presented and compared in the fifth section, followed by conclusions in the last section.

BACKGROUND

A large amount of work has been carried out to improve the quality of multimedia over wireless networks. They can be categorized into two types:

- Studies related to improving multimedia (e.g., video or audio) over wireless networks. They do not consider the mobility of the MN, but attempt to provide a high quality multimedia transmission within the same wireless network for stationary servers and clients.
- Studies related to achieving seamless multimedia transmission during handoffs. They consider mobility of the MH and try to provide a seamless and high quality multimedia transmission when the MH (client) moves from one network to another.

Although our interest in this chapter is seamless multimedia transmission during handoffs, we describe previous work on both categories in the following sections.

Multimedia over Wireless Networks

Ahmed, Mehaoua, and Buridant (2001) worked on improving the quality of MPEG-4 transmis-

sion on wireless using differentiated services (Diffserv). They investigated QoS provisioning between MPEG-4 video application and Diffserv networks. To achieve the best possible QoS, all the components involved in the transmission process must collaborate. For example, the server must use stream properties to describe the QoS requirement for each stream to the network. They propose a solution by distinguishing the video data into important video data and less important video data (such as complementary raw data). Packets which are marked as less important are dropped in the first case if there is any congestion, so that the receiver can regenerate the video with the received important information.

Budagavi and Gibson (2001) improved the performance of video over wireless channels by multiframe video coding. The multiframe coder uses the redundancy that exists across multiple frames in a typical video conferencing sequence so that additional compression can be achieved using their multiframe-block motion compensation (MF-BMC) approach. They modeled the error propagation using the Markov chain, and concluded that use of multiple frames in motion increases the robustness. Their proposed MF-BMC scheme has been shown to be more robust on wireless networks when compared to the base-level H.263 codec which uses single frame-block motion compensation (SF-BMC).

There are a number of studies, such as Stedman, Gharavi, Hanzo, and Steele (1993), Illgner and Lappe (1995), Khansari, Jalai, Dubois, and Mermelstein (1996), and Hanzo and Streit (1995), which concentrate on improving quality of multimedia over wireless networks. Since we are only interested in studies that focus on achieving seamless multimedia transmission during handoff, we do not go into details of studies related to multimedia over wireless networks. Interested readers can use the references given earlier in this paragraph.

Seamless Multimedia over Mobile Networks

Lee, Lee, and Kim (2004) achieved seamless MPEG-4 streaming over a wireless LAN using Mobile IP. They achieved this by implementing packet forwarding with buffering mechanisms in the foreign agent (FA) and performed pre-buffering adjustment in a streaming client. Insufficient pre-buffered data, which is not enough to overcome the discontinuity of data transmission during the handoff period, will result in disruption in playback. Moreover, too much of pre-buffered data wastes memory and delays the starting time of playback. Find the optimal pre-buffering time is, therefore, an important issue in this approach.

Patanapongpibul and Mapp (2003) enable the MH to select the best point of attachment by having all the reachable router advertisements (RA) in a RA cache. RA cache will have the entire router's link whose advertisements are heard by the mobile node. These RAs are arranged in the cache according to a certain priority. The priority is based on two criteria: (1) the link signal strength, that is, signal quality and SNR level, and (2) the time since the RA entry was last updated. So the RAs with highest router priority are forwarded to the IP packet handler for processing. The disadvantage of this method includes extra memory for the RA cache.

Pan, Lee, Kim, and Suda (2004) insert four components in the transport layer of the video server and the client. These four components are: (1) a path management module, (2) a multipath distributor module at the sender, (3) a pair of rate control modules, and (4) a multipath collector module at the receiver. They achieve a seamless video by transferring the video over multiple paths to the destination during handoffs. The overhead of the proposed scheme is two-fold: reduction in transmission efficiency due to transmission of duplicated video packets and transmission of control packets associated with the proposed

scheme, and processing of the proposed scheme at the sender and receiver.

Boukerche, Hong, and Jacob (2003) propose a two-phase handoff scheme to support synchronization of multimedia units (MMU) for wireless clients and distributed multimedia systems. This scheme is proposed for managing MMUs to deliver them to mobile hosts on time. The two-phase scheme consists of: setup handoff and end handoff. In the first phase, setup handoff procedure has two major tasks: updating new arrival BSs and maintaining the synchronization for newly arrived mobile hosts (MHs). If an MH can reach another BS, then MH reports "new BS arrived" to its primary BS. End handoff procedure deals with the ordering of MMUs and with the flow of MMUs for a new MH. Any base station can be a new primary base station. The algorithm notifies MHs, BSs, and servers, and then chooses the closest common node from the current primary base station and new base stations. This method suffers from the disadvantage of additional overhead of updating the base station (BS) with newly arrived BSs and ordering of MMUs.

SIGMA FOR SEAMLESS MULTIMEDIA IN MOBILE NETWORKS

Limitations of previously proposed schemes in achieving seamless multimedia transmission during handoff in a wireless environment have been discussed in the previous section. In this section, we will discuss our proposed handoff scheme, called SIGMA, which has been designed for seamless multimedia transmission during handoffs, followed by its advantages over previous schemes.

Introduction to SIGMA

To aid the reader in getting a better understanding of SIGMA, in this section, we describe the various steps involved in a SIGMA handoff. A detailed description of SIGMA can be found in Fu, Ma, Atiquzzaman, and Lee (2005). We will use the stream control transmission protocol (Stewart, 2005), a new emerging transport layer protocol from IETF, to illustrate SIGMA.

Stream control transmission protocol's (SCTP) multi-homing (see Figure 2) allows an association between two endpoints to span across multiple IP addresses or network interface cards. One of the addresses is designated as the primary while the other can be used as a backup, in the case of failure of the primary address, or when the upper layer application explicitly requests the use of the backup. Retransmission of lost packets can also be done over the secondary address. The built-in support for multi-homed endpoints by SCTP is especially useful in environments that require high-availability of the applications, such as Signaling System 7 (SS7) transport. A multi-homed SCTP association can speedup recovery from link failure situations without interrupting any ongoing data transfer. Figure 2 presents an example of SCTP multi-homing where two nodes, CN and MH, are connected through two wireless networks, with MH being multi-homed. One of MN's IP addresses is assigned as the primary address for use by CN for transmitting data packets; the other IP address can be used as a backup in case of primary address failure.

STEP 1: Obtain New IP Address

Referring to Figure 2, the handoff preparation procedure begins when the MH moves into the overlapping radio coverage area of two adjacent subnets. Once the MH receives the router advertisement from the new access router (AR2), it should initiate the procedure of obtaining a new IP address (IP2 in Figure 2). This can be accomplished through several methods: DHCP, DHCPv6, or IPv6 Stateless Address Autoconfiguration (SAA) (Thomson & Narten, 1998). The main difference between these methods lies in whether the IP ad-

dress is generated by a server (DHCP/DHCPv6) or by the MH itself (IPv6 SAA). For cases where the MH is not concerned about its IP address but only requires the address to be unique and routable, IPv6 SAA is a preferred method for SIGMA to obtain a new address since it significantly reduces the required signaling time.

STEP 2: Add IP Addresses to Association

When the SCTP association is initially setup, only the CN's IP address and the MH's first IP address (IP1) are exchanged between CN and MH. After the MH obtains another IP address (IP2 in STEP 1), MH should bind IP2 into the association (in addition to IP1) and notify CN about the availability of the new IP address (Fu, Ma, Atiquzzaman, & Lee, 2005).

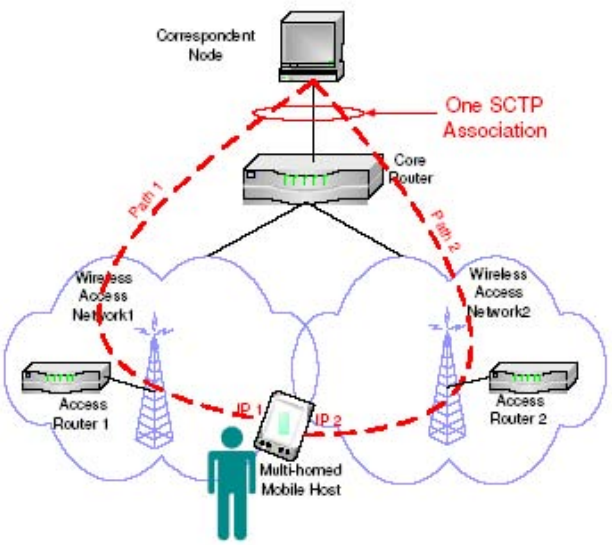
SCTP provides a graceful method to modify an existing association when the MH wishes to notify the CN that a new IP address will be added to the association and the old IP addresses will probably be taken out of the association. The IETF Transport Area Working Group (TSVWG) is working on the "SCTP Address Dynamic Reconfiguration" Internet draft (Stewart, 2005), which defines two new chunk types (ASCONF and ASCONF-ACK) and several parameter types (Add IP Address, Delete IP address, Set Primary Address, etc.). This option will be very useful in mobile environments for supporting service reconfiguration without interrupting on-going data transfers.

In SIGMA, MH notifies CN that IP2 is available for data transmission by sending an ASCONF chunk to CN. On receipt of this chunk, CN will add IP2 to its local control block for the association and reply to MH with an ASCONF-ACK chunk indicating the success of the IP addition. At this time, IP1 and IP2 are both ready for receiving data transmitted from CN to MH.

STEP 3: Redirect Data Packets to New IP Address

When MH moves further into the coverage area of wireless access network2, data path2 becomes increasingly more reliable than data path1. CN can then redirect data traffic to the new IP address

Figure 2. An SCTP association featuring multi-homing



(IP2) to increase the possibility of data being delivered successfully to the MH. This task can be accomplished by the MH sending an ASCONF chunk with the Set-Primary-Address parameter, which results in CN setting its primary destination address to MH as IP2.

STEP 4: Updating the Location Manager

SIGMA supports location management by employing a location manager that maintains a database which records the correspondence between MH's identity and current primary IP address (Reaz, Atiquzzaman, & Fu, 2005). MH can use any unique information as its identity, such as the home address (as in MIP), domain name, or a public key defined in the public key infrastructure (PKI).

Following our example, once the Set-Primary-Address action is completed successfully, MH should update the location manager's relevant entry with the new IP address (IP2). The purpose of this procedure is to ensure that after MH moves from the wireless access network1 into network2, further association setup requests can be routed to MH's new IP address IP2. This update has no impact on existing active associations.

We can observe an important difference between SIGMA and MIP: the location management and data traffic forwarding functions are coupled together in MIP, whereas they are *decoupled in SIGMA to speedup handoff and make the deployment more flexible*.

STEP 5: Delete or Deactivate Obsolete IP Address

When MH moves out of the coverage of wireless access network1, no *new* or *retransmitted* data packets should be directed to address IP1. In SIGMA, MH can notify CN that IP1 is out of service for data transmission by sending an ASCONF chunk to CN (Delete IP Address). Once received, CN will delete IP1 from its local association control block and reply to MH with an ASCONF-ACK chunk indicating the success of the IP deletion.

A less aggressive way to prevent CN from sending data to IP1 is for the MH to advertise a zero receiver window (corresponding to IP1) to CN (Goff, Moronski, Phatak, & Gupta, 2000). This will give CN an impression that the interface (on which IP1 is bound) buffer is full and cannot receive any more data. By deactivating instead of deleting the IP address, SIGMA can adapt more gracefully to MH's zigzag (often referred to as ping pong) movement patterns and reuse the previously obtained IP address (IP1), as long as the lifetime of IP1 has not expired. This will reduce the latency and signaling traffic that would have otherwise been caused by obtaining a new IP address.

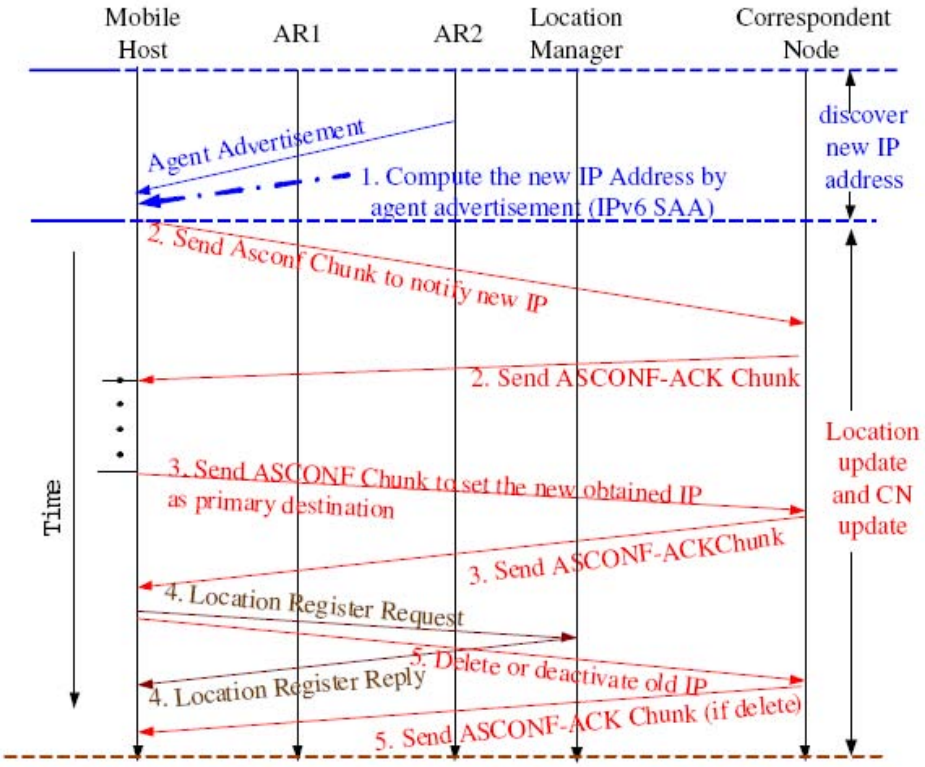
Timing Diagram of SIGMA

Figure 3 summarizes the signaling sequences involved in SIGMA. Here we assume IPv6 SAA and MH initiated Set-Primary-Address. Timing diagrams for other scenarios can be drawn similarly, but are not shown here because of space limitations. In this figure, the numbers before the events correspond to the step numbers in the previous sub-sections, respectively.

Advantages of SIGMA over the Previous Works

A number of previous work have considered seamless multimedia transmission during handoff, as mentioned in the second section, which have their own disadvantages. Here, we discuss the advantages of SIGMA over previous work. Lee et al. (2004) performed pre-buffering adjustment in client. Playback disruption may occur if the pre-buffered data is not enough to overcome the discontinuity of data transmission that occurs during handoff. Moreover, excessive pre-buffered

Figure 3. Timeline of signaling in SIGMA



data wastes memory usage and delays the starting time of playback. Find the optimal pre-buffering time is an important issue in this approach. Since SIGMA does not pre-buffer any data in the client, such optimization issues are not present in SIGMA.

Patanapongpibul et al. (2003) use the router advertisement (RA) cache. The disadvantage of this method is that it needs extra memory for RA cache; SIGMA does not involve any caching and hence does not suffer from such memory problems. Pan et al. (2004) use multipath (as discussed earlier), which suffers from (1) reduction in bandwidth efficiency due to transmission of duplicated video packets and transmission of control packets associated with the proposed scheme, and (2) processing overhead at the sender and receiver. Absence of multipaths or duplicate video packets in SIGMA results in higher link bandwidth efficiency.

Boukerche et al. (2003) proposed a two-phase handoff scheme which has additional overhead of updating the base station (BS) with newly arrived BSs, and also ordering of multimedia units (MMUs). In SIGMA, there is no feedback from MH to any of the base stations, and hence does not require ordering of multimedia units or packets.

EXPERIMENTAL TESTBED

Having reviewed the advantages of SIGMA over other schemes for multimedia transmission in the previous section, in this section, we present experimental results for SIGMA as obtained from an experimental setup we have developed

at the University of Oklahoma. We compare the results of handoff performance during multimedia transmission over both SIGMA and Mobile IP. To make a fair comparison, we have used the same test bed for both MIP and SIGMA. Figure 4 (to be described later) shows the topology of our test bed, which has been used by a number of researchers—Seol, Kim, Yu, and Lee (2002), Wu, Banerjee, Basu, and Das (2003), Onoe, Atsumi, Sato, and Mizuno (2001)—for measurement of handoff performance. The difference in data communication between the CN and the MH for MIP and SIGMA lies in the lower layer sockets: the file sender for MIP is based on the regular TCP socket, while that for SIGMA is based on SCTP socket. We did not use the traditional *ftp* program for file transfer because it was not available for the SCTP protocol. To obtain access to the SCTP socket, we used Linux 2.6.2 kernel with Linux Kernel SCTP (LKSCTP) version 2.6.2-0.9.0 on both CN and MN. A number of MIP implementations, such as HUT Dynamics (HUT), Stanford Mosquito (MNET), and NUS Mobile IP (MIP), are publicly available. We chose HUT Dynamics for testing MIP in our test bed due to the following reasons: (1) Unlike Stanford Mosquito, which integrates the FA and MN, HUT Dynamics implements HA, FA, and MH daemons separately. This architecture is similar to SIGMA where the two access points and MH are separate entities. (2) HUT Dynamics implements hierarchical FAs, which will allow future comparison between SIGMA and hierarchical Mobile IP. Our MIP testbed consists four nodes: correspondent node (CN), foreign agent (FA), home agent (HA), and mobile node (MN). All the nodes run corresponding agents developed by HUT Dynamics. The hardware and software configuration of the nodes are given in Table 1.

The CN and the machines running the HA and FA are connected to the Computer Science (CS) network of the University of Oklahoma, while the MH and access points are connected to two separate private networks. The various IP addresses are shown in Table 2. IEEE 802.11b is used to connect the MH to the access points.

The network topology of SIGMA is similar to the one of Mobile IP except that there is no HA or FA in SIGMA. As shown in Figure 4, the machines which run the HA and FA in the case of MIP act as gateways in the case of SIGMA. Table 1 shows the hardware and software configuration for the SIGMA experiment. The various IP addresses are shown in Table 2. The experimental procedure of Mobile IP and SIGMA is given next:

Table 1. Mobile IP and SIGMA testbed configurations

Node	**Hardware**	**Software**	**Operating System**
Home Agent(MIP) Gateway1 (SIGMA)	Desktop, two NICs	HUT Dynamics 0.8.1 Home Agent Daemon (MIP)	Redhat Linux 9 kernel 2.4.20
Foreign Agent (MIP) Gateway2 (SIGMA)	Desktop, two NICs	HUT Dynamics 0.8.1 Foreign Agent Daemon (MIP)	Redhat Linux 9 kernel 2.4.20
Mobile Node	Dell Inspiron- 1100 Laptop, one Avaya 802.11b wireless card	HUT Dynamics 0.8.1 Mobile Node Daemon (MIP), File receiver	Redhat Linux 9 kernel 2.4.20
Correspondent Node	Desktop, one NIC	File sender	Redhat Linux 9 2.6.20

Table 2. Mobile IP and SIGMA network configurations

Node	Network Configuration
Home Agent (MIP) Gateway1 (SIGMA)	eth0: 129.15.78.171, gateway 129.15.78.172; eth1:10.1.8.1
Foreign Agent (MIP) Gateway2 (SIGMA)	eth0: 129.15.78.172 gateway 129.15.78.171; eth1: 10.1.6.1
Mobile Node	Mobile IP's Home Address: 10.1.8.5 SIGMA's IP1: 10.1.8.100 SIGMA's IP2 : 10.1.6.100
Correspondent Node	129.15.78.150

1. Start with the MH in Domain 1.
2. **For Mobile IP:** Run HUT Dynamics daemons for HA, FA, and MN. **For SIGMA:** Run the SIGMA handoff program, which has two functions: (1) monitoring the link layer signal strength to determine the time to handoff, and (2) carrying out the signaling shown in Figure 4.
3. Run file sender/video server and file receiver/video client (using TCP sockets for Mobile IP, using SCTP sockets for SIGMA) on CN and MN, respectively.
4. Run Ethereal (ETHEREAL) on the CN and MH to capture packets.
5. Move MH from Domain 1 to Domain 2 to perform handoff by Mobile IP and SIGMA. Capture all packets sent from CN and received at MN.

RESULTS

Various results were collected on the experimental setup and procedure described earlier. In this section, we present two kinds of results: file transfer and multimedia transmission. The reason for showing the results of file transfer is to prove that SIGMA achieves seamless handoff not only for multimedia but also for file transfers.

Results for File Transfer

In this section, we present and compare the results of handoffs using MIP and SIGMA for file transfer. For comparison, we use throughput, RTT, and handoff latency as the performance measures. *Throughput* is measured by the rate at which packets are received at the MN. *RTT* is the time required for a data packet to travel from the source to the destination and back. We define *handoff latency* as the time interval between the MH receiving the last packet from Domain 1 (previous network) and the first packet from Domain 2 (the new network). The experimental results are described next.

Results from Mobile IP Handoff

Figure 5 shows the throughput during Mobile IP handoff between Domain 1 and Domain 2. The variations in throughput within HA (from 20 second to 30 second) and within FA (from 37 second to 60 second) are due to network congestion arising from cross traffic in the production CS network.

The average throughput before, during and after handoff are 2.436 Mbps, 0 Mbps and 2.390 Mbps, respectively. Figure 6 shows the packet trace during MIP handoff. The actual handoff latency for MIP can be clearly calculated by having a zoomed-in view of the packet trace graph. Figure 7 shows a zoomed-in view of the packet trace, where the calculated handoff latency is eight seconds for Mobile IP. Figure 8 shows the RTT for the MIP handoff. As we can see, the RTT is high for eight seconds (the handoff latency time), during the handoff.

Figure 4. SIGMA and Mobile IP testbed

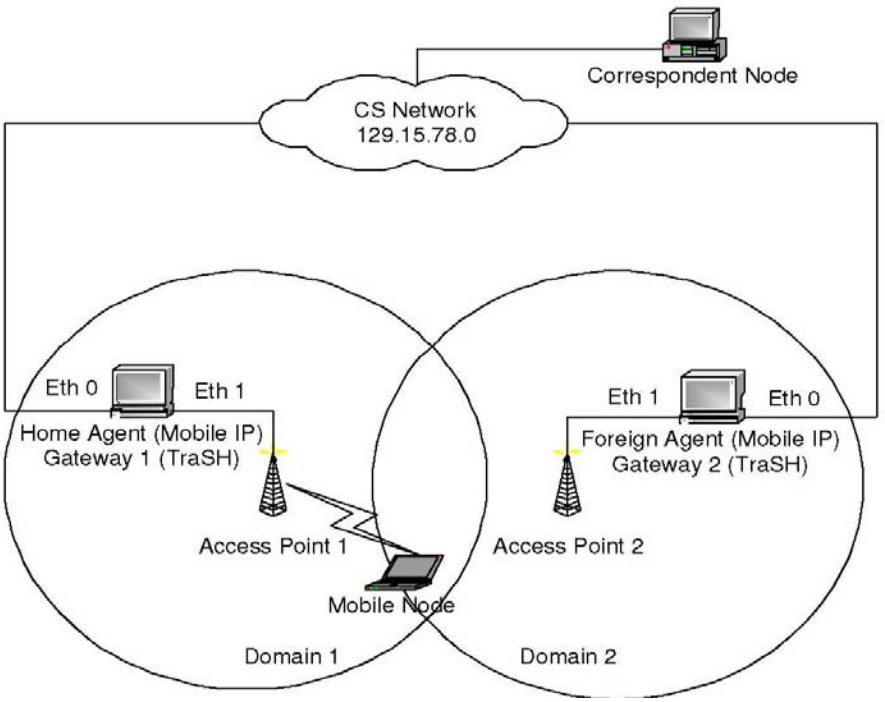


Figure 5. Throughput during MIP handoff

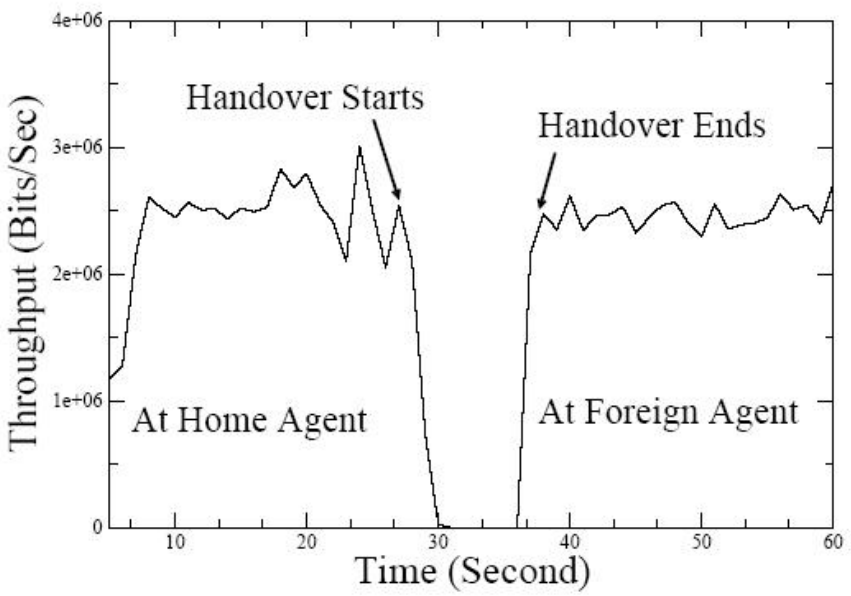


Figure 6. Packet trace during MIP handoff

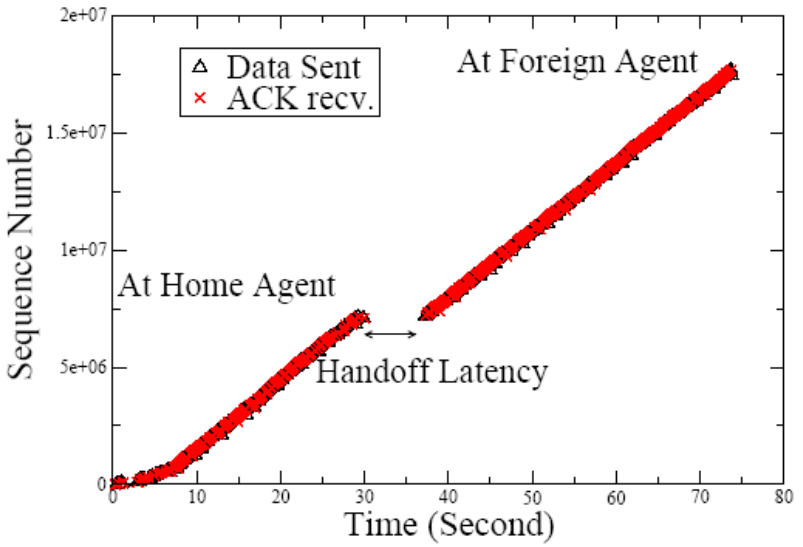
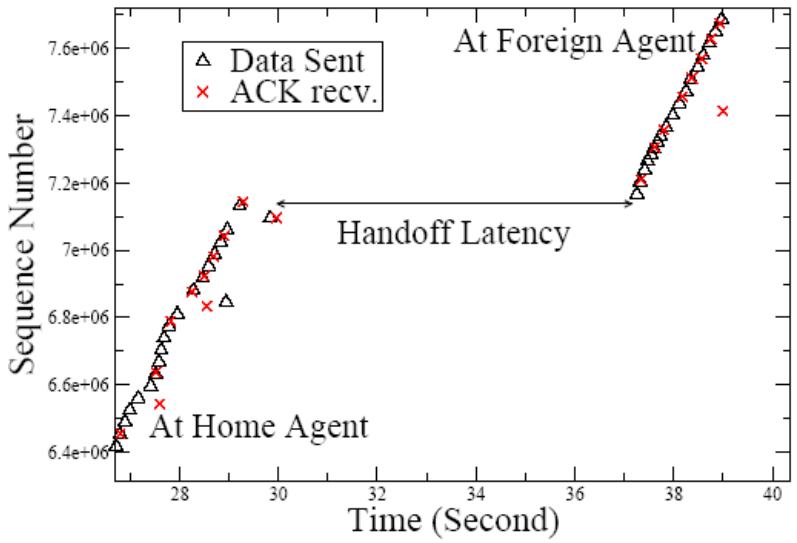
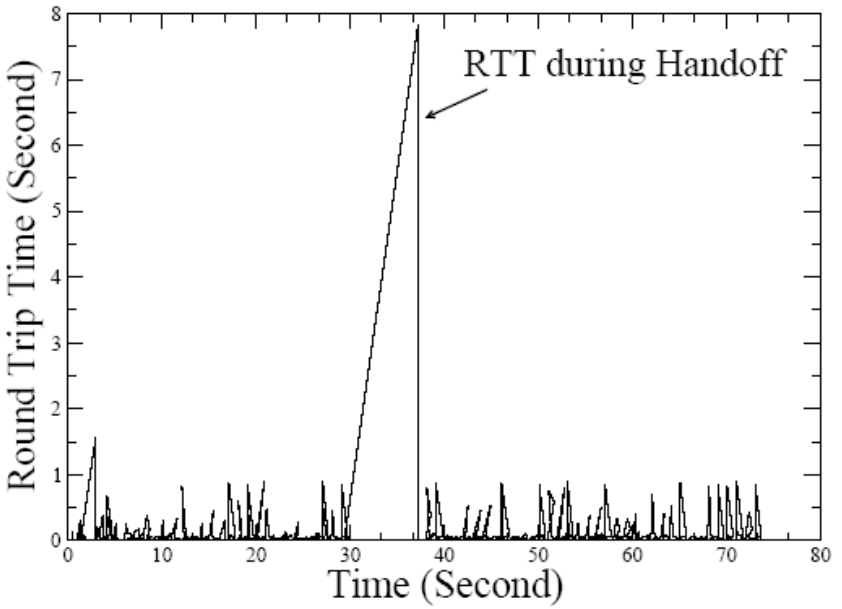


Figure 7. Zoomed in view during MIP handoff instant



The registration time (or registration latency) is also a part of the handoff latency. Registration latency, the time taken by the MH to register with the agent (HA or FA), is calculated as follows. Ethereal capture showed that the MH sent a registration request to the HA at time t = 14.5123 second and received a reply from the HA at t = 14.5180 second. Hence, the calculated registration time for registering with HA is 5.7 milliseconds. Similarly, during MIP handoff, Ethereal capture showed that the MH sent a registration request to FA at time t =7.1190 second and received a reply from the FA at t =7.2374, resulting in a registration time of 38.3 milliseconds. This is due to the fact that after the MH registers with the HA, it can directly register with the HA. On the other hand, if it registers with the FA, the MH registers each new care-of-address with its HA possibly

Figure 8. RTT during MIP handoff



through FA. The registration latency is, therefore, higher when the MH is in the FA.

Results from SIGMA Handoff

Figure 9 shows the throughput during SIGMA handoff where it can observed that the throughput does not go to zero. The variation in throughput is due to network congestion arising from cross traffic in the production CS network. Although we cannot see the handoff due to it being very small, it should be emphasized that the ethereal capture showed the handoff starting and ending at t = 60.755 and t = 60.761 seconds, respectively, that is, a handoff latency of six milliseconds.

Figure 10 shows the packet trace during SIGMA handoff. It can be seen that packets arrive at the MH without any gap or disruption; this is also a powerful proof of SIGMA's smoother handoff as compared to handoff in Mobile IP. This experimentally demonstrates that *a seamless handoff can be realized with SIGMA.* Figure 11 shows a zoomed-in view of the packet trace during the SIGMA handoff period; a handoff latency of six milliseconds can be seen between the packets arriving at the old and new paths.

Figure 12 shows the RTT during SIGMA handoff. A seamless handoff is evident from the absence of any sudden RTT increase during handoff.

Result of Multimedia Data Transfer

To test the handoff performance for multimedia over SIGMA, we used a streaming video client and a streaming server at the MH and CN, respectively (details in the fourth section). Apple's Darwin Streaming Server (DARWIN) and CISCO's MPEG4IP player (MPEG) were modified to stream data over SCTP. A seamless handoff, with no interruption in the video stream, was achieved with SIGMA.

Figure 13 shows the throughput of multimedia (video) data, when the MH moves between subnets. The connection request and setup between the client and server is carried out during the first 10 seconds. It can be seen that the throughput does not drop during handoff at time = 31 second

Figure 9. Throughput during SIGMA handoff

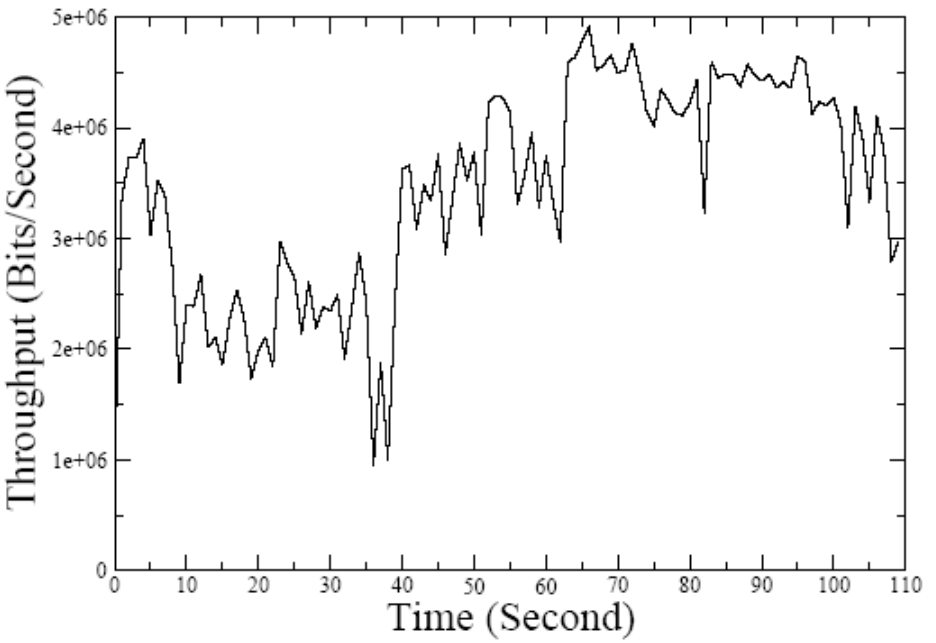


Figure 10. Packet trace during SIGMA handoff

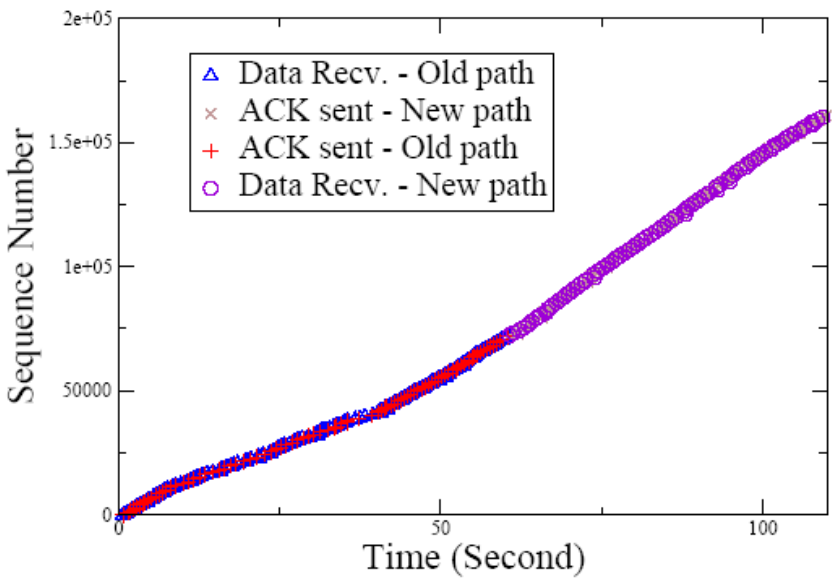


Figure 11. Zoomed in view during SIGMA handoff

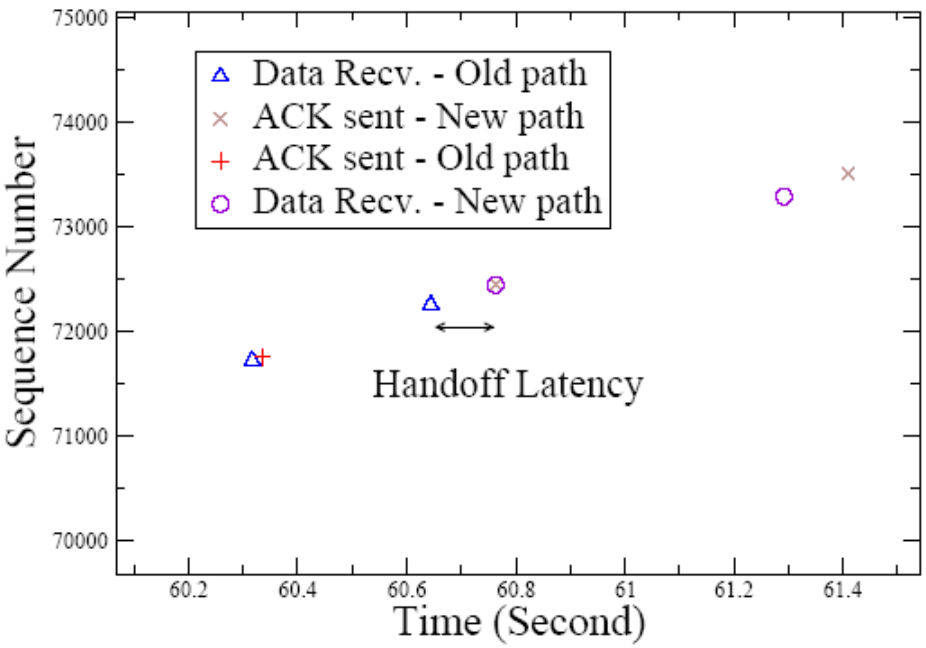


Figure 12. RTT during SIGMA handoff

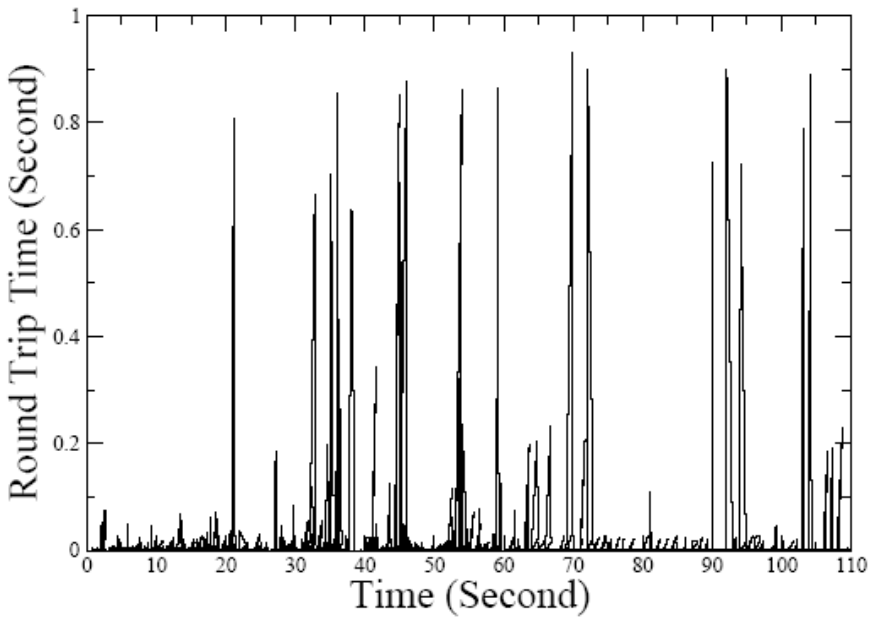
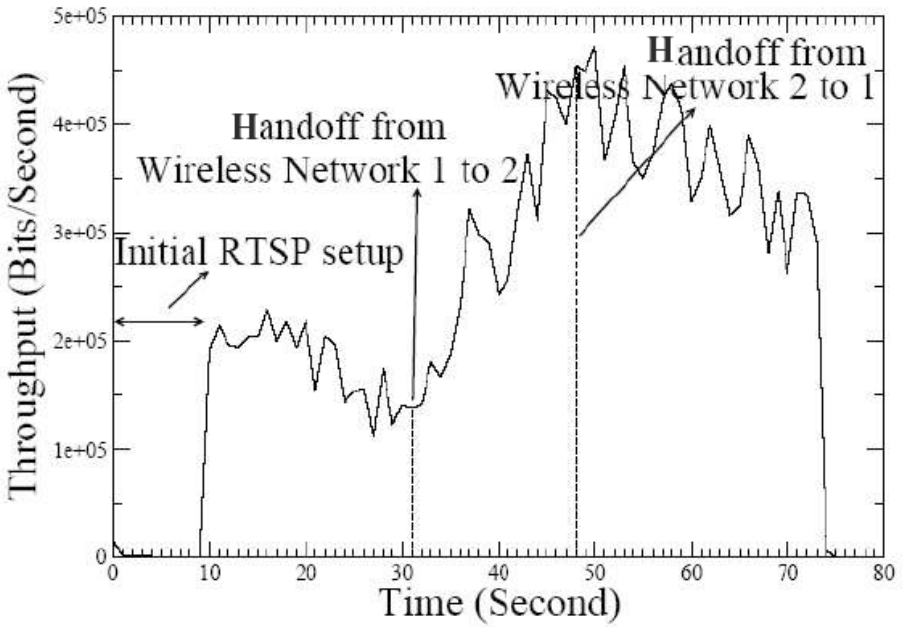


Figure 13. Throughput of video during SIGMA handoff



when MH moves from wireless network 1 to 2. A second handoff takes place when the MH moves from network 2 to network 1 at time = 48. It is seen that seamless handoff is achieved by SIGMA for both the handoffs.

Figure 14 shows a screen capture of the MPEG4IP player used in our experiment. Figure 15 shows the video playing in the player during handoff, where "rtsp://129.15.78.139/fta.sdp" represents the server's IP address and the streaming format (SDP).

Comparison of SIGMA and MIP Handoffs

We observed previously that the registration time of MIP was only 0.1 second, and the handoff latencies of MIP and SIGMA were eight seconds and six milliseconds, respectively. We describe the reasons for the MIP handoff latency being much longer than its registration time in the following:

1. In HUT Dynamics, the MIP implementation used in this study, the MH obtains a registration lifetime after every successful registration. It originates another registration on expiry of this lifetime. So it is possible for the MH to postpone registration even after it has completed a link layer handoff and received FA advertisements. This may introduce some delay which can be up to the duration of a life time.
2. As mentioned in the previous section, the registration of MH also costs some time, measured as 38.3 milliseconds in our testbed.

The handoff latency in MIP comes from three factors: (1) remaining home registration lifetime after link layer handoff which can be from zero to a lifetime, (2) FA advertisement interval plus the time span of last time advertisement which is not listened by MN, and (3) registration latency. During these three times, the CN cannot communicate through either the previous path because it

Figure 14. Screen shot of MPEG4-IP player

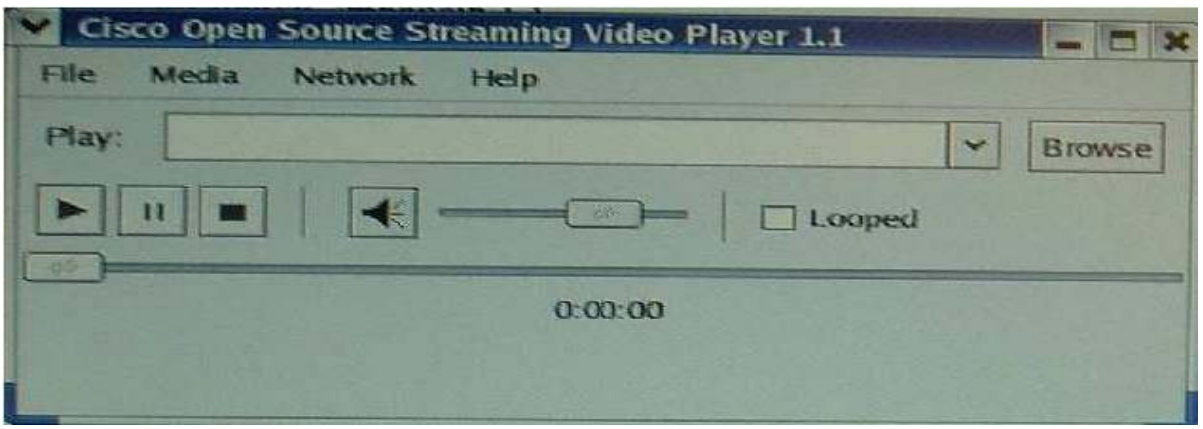


Figure 15. Screen-shot of MPEG4-IP player playing streaming video

has completed link layer handoff, or the new path because MH has not yet completed the registration. As a result, the throughput was zero during this time. Obviously, such shortcoming has been eliminated in SIGMA through multi-homing and decoupling of registration and data transfer. Consequently, data continue to flow between the CN and MH during the handoff process.

CONCLUSION AND FUTURE TRENDS

We have shown that SIGMA achieves seamless multimedia transmission during handoff between wireless networks. As future work, video streaming can be tested over SIGMA during vertical handoffs, that is, between wireless LANs, cellular, and satellite networks.

ACKNOWLEDGMENT

The work reported in this chapter was funded by National Aeronautics and Space Administration (NASA) grant no. NAG3-2922.

REFERENCES

Ahmed, T., Mehaoua, A., & Buridant, G. (2001). Implementing MPEG-4 video on demand over IP differentiated services. *Global Telecommunications Conference, GLOBECOM*, San Antonio, TX, November 25-29 (pp. 2489-2493). Piscataway, NJ: IEEE.

Boukerche, A., Hong, S., & Jacob, T., (2003). A two-phase handoff management scheme for synchronizing multimedia units over wireless networks. *Proc. Eighth IEEE International Symposium on Computers and Communication*, Antalya, Turkey, June-July (pp. 1078-1084). Los Alamitos, CA: IEEE Computer Society.

Budagavi, M., & Gibson, J. D. (2001, February). Multiframe video coding for improved performance over wireless channels. *IEEE Transactions on Image Processing, 10*(2), 252-265.

DARWIN. Retrieved June 23, 2005, from http://developer.apple.com/darwin/projects/streaming/

ETHEREAL. Retrieved June 30, 2005, from www.ethereal.com

Fu, S., Atiquzzaman, M., Ma, L., & Lee, Y. (2005, November). Signaling cost and performance of SIGMA: A seamless handover scheme for data networks. *Journal of Wireless Communications and Mobile Computing, 5*(7), 825-845.

Fu, S., Ma, L., Atiquzzaman, M., & Lee, Y. (2005). Architecture and performance of SIGMA: A seamless mobility architecture for data networks. *40th IEEE International Conference on Communications (ICC)*, Seoul, Korea, May 16-20 (pp. 3249-3253). Institute of Electrical and Electronics Engineers Inc.

Goff, T., Moronski, J., Phatak, D. S., & Gupta, V. (2000). Freeze-TCP: A true end-to-end TCP enhancement mechanism for mobile environments. *IEEE INFOCOM*, Tel Aviv, Israel, March 26-30 (pp. 1537-1545). NY: IEEE.

Hanzo, L., & Streit, J. (1995, August). Adaptive low-rate wireless videophone schemes. *IEEE Trans. Circuits Syst. Video Technol., 5*(4), 305-318.

HUT. Retrieved June 1, 2005, from http://www.cs.hut.fi/research/dynamics/

Illgner, R., & Lappe, D. (1995). Mobile multimedia communications in a universal telecommunications network. *Proc. SPIE Conf. Visual Communication Image Processing*, Taipei, Taiwan, May 23-26 (pp. 1034-1043). USA: SPIE.

Khansari, M., Jalai, A., Dubois, E., & Mermelstein, P. (1996, February). Low bit-rate video transmission over fading channels for wireless microcellular system. *IEEE Trans. Circuits Syst. Video Technol., 6*(1), 1-11.

Lee, C. H., Lee, D., & Kim, J. W. (2004). Seamless MPEG-4 video streaming over Mobile-IP enabled wireless LAN. *Proceedings of SPIE, Multimedia Systems and Applications*, Philadelphia, Pennsylvania, October (pp. 111-119). USA: SPIE.

LKSCTP. Retrieved June 1, 2005, from http://lksctp.sourceforge.net

MIP. Retrieved June 1, 2005, from opensource.nus.edu.sg/projects/mobileip/mip.html

MNET. Retrieved June 1, 2005, from http://mosquitonet.stanford.edu/

MPEG. Retrieved June 1, 2005, from http://mpeg4ip.sourceforge.net/faq/index.php

Onoe, Y., Atsumi, Y., Sato, F., & Mizuno, T. (2001). A dynamic delayed ack control scheme on Mobile IP networks. *International Conference on Computer Networks and Mobile Computing*, Los Alamitos, CA, October 16-19 (pp. 35-40). Los Alamitos, CA: IEEE Computer Society.

Pan, Y., Lee, M., Kim, J. B., & Suda, T. (2004, May). An end-to-end multipath smooth handoff scheme for streaming media. *IEEE Journal on Selected Areas in Communications, 22*(4), 653-663.

Patanapongpibul, L., & Mapp, G. (2003). A client-based handoff mechanism for Mobile IPv6 wireless networks. *Proc. Eighth IEEE International Symposium on Computers and Communications*, Antalya, Turkey, June-July (pp. 563-568). Los Alamitos, CA: IEEE Computer Society.

Perkins, C. (1996). IP mobility support. *IETF RFC 2002*, October.

Reaz, A. S., Atiquzzaman, M., & Fu, S. (2005). Performance of DNS as location manager. *IEEE Globecom*, St. Louis, MO, November 28-December 2 (pp. 359-363). USA: IEEE Computer Society.

Seol, S., Kim, M., Yu, C., & Lee., J. H. (2002). Experiments and analysis of voice over MobileIP. *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Lisboa, Portugal, September 15-18 (pp. 977-981). Piscataway, NJ: IEEE.

Stedman, R., Gharavi, H., Hanzo, L., & Steele, R. (1993, February). Transmission of subband-coded images via mobile channels. *IEEE Trans. Circuit Syst. Video Technol., 3*, 15-27.

Stewart, R. (2005, June). *Stream control transmission protocol (SCTP) dynamic address configuration.* IETF DRAFT, draft-ietf-tsvwgaddip-sctp-12.txt.

Thomson, S., & Narten, T. (1998, December). *IPv6 stateless address autoconfiguration.* IETF RFC 2462.

Wu, W., Banerjee, N., Basu, K., & Das, S. K. (2003). Network assisted IP mobility support in wireless LANs. *Second IEEE International Symposium on Network Computing and Applications, NCA'03*, Cambridge, MA, April 16-18 (pp. 257-264). Los Alamitos, CA: IEEE Computer Society.

This work was previously published in Mobile Multimedia Communications: Concepts, Applications, and Challenges, edited by G. Karmakar, pp. 24-44, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.6
Mobile Multimedia:
Communication Technologies, Business Drivers, Service and Applications

Ismail Khalil Ibrahim
Johannes Kepler University Linz, Austria

Ashraf Ahmad
National Chiao Tung University, Taiwan

David Taniar
Monash University, Australia

ABSTRACT

Mobile multimedia, referring to multimedia information exchange over wireless networks or wireless Internet, is made possible due to the popularity and evolution of mobile computing devices, coupled with fast and affordable mobile networks. This chapter discusses various state-of-the-art communication technologies to support mobile multimedia. The range of complexity of applications and services provided to end-users also play an important part in the success of mobile multimedia.

INTRODUCTION

Number of subscribers for mobile communications has increased much faster than predicted, particularly for terrestrial use. In the year 2000, number of mobile subscribers was approximately 400 million worldwide, and in the year 2010 more than 1.8 billion mobile subscribers are estimated. An interesting fact was presented in a new report by Telecommunications Management Group, Inc. (TMG) (Wireless World Forum) providing the statistical basis to show the number of mobile multimedia users exceeding 100 million in 2004. This breathtaking fact inspires us to start

researching the mobile multimedia in all possible related aspects. In order to provide experts and researcher in the field of mobile multimedia, a description of basic definition of mobile multimedia is introduced, and then essential business driver controlling the mobile multimedia is illustrated. A full and up-to-date description of technologies beneath the mobile multimedia is presented. In addition to the services and applications, a set of mobile multimedia is discussed thoroughly, as well as anticipating the future of mobile multimedia.

The demand for mobile access to data no matter where the data is stored, what is kind of data and where the user happens to be, in addition to the explosive growth of the Internet and the rising popularity of mobile devices are among the factors that have created a dynamic business environment, where both handset manufacturers and service provider operators companies are competing to provide customers access to information resources and services anytime, anywhere.

Advances in wireless networking, specifically the development of the IEEE 802.11 and IEEE 802.16 protocol family and the rapid deployment and growth of GSM (and GPRS, 3G, 3GPP), have enabled a broad spectrum of novel and out breaking solutions for new applications and services. Voice services are no longer sufficient to satisfy customers' business and personal requirements. More and more people and companies are demanding for mobile access to multimedia services. Mobile multimedia seems to be the next mass market in mobile communications following the success of GSM and SMS. It enables the industry to create products and services to meet the consumer needs better. However, an innovation itself does not guarantee a success; it is necessary to be able to predict the new technology adaptation behavior and to try to fulfill customer needs rather than to wait for a demand pattern to surface.

The major step from the second generation to third generation and further to fourth generation was the ability to support advanced and wideband multimedia services, including emails, file transfers, and distribution services like radio, TV, and software provisioning (e.g., software download). These multimedia services can be symmetrical and asymmetrical services, real-time, and non real-time services.

It is beyond all expectations that mobile multimedia will create significantly added values for costumers by providing mobile access to Internet-based multimedia services, video conferencing, and streaming. Mobile multimedia is one of the mainstream systems for the next generation mobile communications, featuring large voice capacity, multimedia applications and high-speed mobile data services (Bull, Canagarajah, & Nix, 1999). As for the technology, the trend in the radio frequency area is to shift from narrowband to wideband with a family of standards tailored to a variety of application needs. Many enabling technologies including WCDMA, software-defined radio, intelligent antennas, and digital processing devices are greatly improving the spectral efficiency of third generation systems. In the mobile network area, the trend is to move from traditional circuit-switched systems to packet-switched programmable networks that integrate both voice and packet services, and eventually evolve towards an all-IP network (Bi, Zysman, & Menkes, 2001).

While for the information explosion, the addition of mobility to data communications systems has enabled a new generation of services not meaningful in a fixed network, such as positioning-based services. However, the development of mobile multimedia services has only started and in the future we will see new application areas opening up (Bi, Zysman, & Menkes, 2001; Blair, Coulson, Davies, Robin, & Fitzpatrick, 1997; Brown & Syfrig, 1996; Bruno, Conti, & Gregori, 2001; Bull, Canagarajah, & Nix, 1999).

Research in mobile multimedia is typically focused on bridging the gap between the high resource demands of multimedia applications and the limited bandwidth and capabilities offered

by state-of-the-art networking technologies and mobile devices.

Communication engineering approaches this problem by considering not only characteristics of the networks and devices used, but also on the tasks and objectives the user is pursuing when applying/demanding mobile multimedia services and exploit this information to better adapt those services to the users' needs. This method is referred to it as user centric multimedia processing approach.

Mobile Multimedia

External market studies have predicted that in Europe in the year 2010 more than 90 million mobile subscribers will use mobile multimedia services and will generate about 60% of the traffic in terms of transmitted bits. In China, the DGI predicted that there will be over 500 million mobile phones in China by year 2008, and over 150 million for multimedia applications. These results grab our attention how important it is to precisely define the mobile multimedia and its related terms.

Mobile multimedia can be defined as a set of protocols and standards for multimedia information exchange over wireless networks. It enables information systems to process and transmit multimedia data to provide end users with services from various areas, such as the mobile working place, mobile entertainment, mobile information retrieval, and context-based services.

Multimedia information as combined information presented by more than one media type (i.e. text, pictures, graphics, sounds, animations, videos) enriches the quality of the information and is a way to represent reality as adequately as possible. Multimedia allows users to enhance their understanding of the provided information and increases the potential of person to person and person to system communication.

Mobility as one of the key drivers of mobile multimedia can be decomposed into:

1. **User mobility:** The user is forced to move from one location to location during fulfilling his activities. For the user, the access to information and computing resources is necessary regardless his actual position (e.g., terminal services, VPNs to company-intern information systems).
2. **Device mobility:** User activities require a device to fulfill his needs regardless of the location in a mobile environment (e.g., PDAs, notebooks, tablet pc, cell-phones, etc.).
3. **Service mobility:** The service itself is mobile and can be used in different systems and can be moved seamlessly among those systems (e.g., mobile agents).

The special requirements coming along with the mobility of users, devices, and services; and specifically the requirements of multimedia as traffic type; bring the need of new paradigms in software-engineering and system-development; but also in non-technical issues such as the emergence of new business models and concerns about privacy, security, or digital inclusion, to name a few.

For instance, in the context of mobile multimedia, 3G communication protocols have great deals. Even some mobile protocol experts tend to define 3G as a mobile multimedia, personal services, the convergence of digitalization, mobility, and the internet, new technologies based on global standards, the entire of the aforementioned terms. In 3G, the end user will be able to access the mobile internet at the bandwidth at various bit rates. This makes great challenges for handset device manufacturers and mobile network operators. In addition, a large number of application and related issues need to be addressed considering the heterogeneity nature of the Internet. Also, the various and rich content of the internet should be considered whenever one of the roles of 3G is being deployed.

In network traffic point of view, the majority of traffic is changing from speech-oriented com-

munications to multimedia communications. It is also generally expected that due to the dominating role of mobile wireless access, the number of portable handsets will exceed the number of PCs connected to the Internet. Therefore, mobile terminals will be the major person-machine interface in the future instead of the PC. Due to the dominating role of IP-based data traffic in the future, the networks and systems have to be designed for economic packet data transfer. The expected new data services are highly bandwidth consuming. This results in higher data rate requirements for future systems.

Business Drivers

The key feature of mobile multimedia is reaching customers and partners, regardless of their locations and delivering multimedia content to the right place at the right time. Key drivers of this technology are on the one hand technical and on the other business drivers.

Evolutions in technology pushed the penetration of the mobile multimedia market and made services in this field feasible. The miniaturization of devices and the coverage of radio networks are the key technical drivers in the field of mobile multimedia.

1. **Miniaturization:** The first mobile phones had brick-like dimensions. Their limited battery capacity and transmission range restricted their usage in mobile environments. Actual mobile devices with multiple features fit into cases with minimal dimensions and can be (and are) carried by the user in every situation.
2. **Vehicle manufacturer:** Furthermore, mobility also calls for new types of services (and thus revenues). Vehicle manufacturers want to improve the ears' man-machine interface by using superior input/output devices. An open application platform would allow upgrading of multimedia equipment during the lifecycle of a vehicle, which is much longer than the lifetime of computing equipment. Safety equipment for automated emergency and breakdown calls brings positioning hardware into the car and thus enabling other location-aware services. But vehicle makers also seek an after-market relationship to their customers: Once having ears connected, they can offer ear-specific services, including direction-finding and safeguarding support.
3. **Radio networks:** Today's technology allows radio networks of every size for every application scenario. Nowadays public wireless wide-area networks cover the bulk of areas especially in congested areas. They enable (most of the time) adequate quality of service. They allow location-independent service provision and virtual private network access.
4. **Mobile terminal manufacturers:** Mobile terminal manufacturers serve individual people instead of households. Since there are more individuals than households, the market is naturally bigger then the home terminal market. Furthermore, there is a large potential for use and fashion-based diversification and innovation of terminals.
5. **Market evolution:** The market for mobile devices changed in the last years. Ten years ago the devices were not really mobile (short-time battery operation, heavy, and large devices) but therefore they have been expensive and affordable just for high-class business people. Shrinking devices and falling operation (network) costs made mobile devices to a mass-consumer-good available and affordable for everyone. The result is dramatic subscriber growth and therefore a new increasing market for mobile multimedia services.
6. **Subscribers:** Persons spend a good percentage of their lifetime traveling, either for business or leisure, while they want to

stay connected in every respect. This desire is more than proven by the current sales figures for mobile phones and the emerging standards for mobile narrow-band data services.

7. **Service evolution:** The permanent increasing market brought more and more sophisticated services, starting in the field of telecommunication from poor quality speech-communication to real-time video conferencing. Meanwhile, mobile multimedia services provide rich media content and intelligent context-based services.
8. **Vehicle terminal manufacturers:** Vehicle terminal manufacturers currently suffer from vertical markets due to high customization efforts for OEM products. An open application platform would help them to reduce development time and costs. It is also a key driver for after market products. A wide range of services will increase the number of terminals sold.
9. **Ears:** For ear drivers, security and travel assistance are important aspects as well. They probably want to use the same services in the car they are used to at home and in the office. This is only possible with an open application platform.

Technology drives mobile multimedia with new means to communicate, cache, process, and display multimedia content:

- **Connectivity:** New means to communicate enable new services to be provided.
- **Memory and persistent storage:** Developing memory technology allows caching of more content and offline processing, thus creating the illusion of instant access to interactive remote content. For example, audio/video content and whole Websites may be downloaded in background and consumed offline. This is above all important for data broadcast services, which are transmitted in different fashions.
- **Processing:** More processing resources with less power consumption allow rendering of more complex multimedia content.
- **Display:** Visualizing multimedia content demands for cheap high resolution displays that comply to "handset" requirements "screen driver, pixel per bits, width, height."

The value chain of mobile multimedia services describes the players involved in the business with mobile multimedia. Every service in the field of mobile multimedia requires that their output and service fees must be divided to them considering interdependencies in the complete service life-cycle.

1. **Network operators:** They provide end-users with the infrastructure to access services mobile via wireless networks (e.g., via GSM/GPRS/UMTS). The network operators want to boost the sales of network bandwidth, by enabling new types of services with new network technologies. In many countries a close cooperation between the transmitter and cellular network operators has been established to be able to offer hybrid network capacity. Service and content providers see the opportunity to promote and sell their services to people everywhere and anytime, thus increasing the total usage of their services.
2. **Content provider:** Content provider and aggregators license content and prepare it for end-users. They collect information and services to provide customers with convenient service collection adapted for mobile use. In another hand, some national broadcasters are forced by law to provide nationwide TV coverage through terrestrial transmission. They are interested in digital TV, because, firstly, it strongly reduces their transmission costs per channel by high ratio and, secondly, they improve the attractiveness of terrestrial reception which decreased strongly since the beginning of cable and satellite services

3. **Fixed Internet company:** Those companies create the multimedia content. Usually they provide it already via the fixed Internet but are not specialized on mobile service provisioning. They handle the computing infrastructure and content creation.
4. **App developers and device manufacturers:** They deliver hard- and software for mobile multimedia services and are not involved with any type of content creation and delivering.

COMMUNICATION TECHNOLOGIES

Wireless Wide Area Networks

After the first-generation analog mobile systems, the second-generation (2G) mobile digital systems were introduced around 1991 offering higher capacity and lower costs for network operators, while for the users, they offered short messages and low-rate data services added to speech services. Reader may refer to figure one as it holds general comparison among the first to four generation communication system. Presently, the 2G systems are GSM, TDMA, PDC, and cdmaOne. GSM is used in most parts of the world except in Japan, were PDC is the second-generation system used (Dixit, Guo, & Antoniou, 2001).

An important evolution of the 2G systems, sometimes known as 2.5G, is the ability to use packet-switched solution in GPRS (general packet radio system). The main investment for the operators lies in the new packet-switched core network, while the extensions in the radio access network mainly is software upgrades. For the users GPRS offers the possibility to always be online and only pay for the data actually transferred. Data rates of up to 20 kbps per used time slot will be offered, and with multiple time-slots per user in the downlink, attractive services can be offered (Stallings, 2001).

The shift to third-generation in the radio access networks is demanding a lot of efforts. The ITU efforts through IMT-2000 have led to a number of recommendations. These recommendations address areas such as user bandwidth, richness of service offerings (multimedia services), and flexibility (networks that can support small or large numbers of subscribers). The recommendations also specify that IMT-2000 should operate in the 2-GHz band. In general, however, the ITU recommendations are mainly a set of requirements and do not specify the detailed technical solutions to meet the requirements (UMTS Forum, 1998, 2000).

To address the technical solutions, the ITU has solicited technical proposals from interested organizations, and then selected/approved some of those proposals. In 1998, numerous air interface technical proposals were submitted. These were reviewed by the ITU, which in 1999 selected five technologies for terrestrial service (non-satellite based). The five technologies are (Collins & Smith, 2001):

- Wideband CDMA (WCDMA)
- CDMA 2000 (an evolution of IS-95 CDMA)
- TD-SCDMA (time division-synchronous CDMA)
- DECT
- UWC-136 (an evolution of IS-136)

Here is a brief description of each one of these selected technologies.

Wideband CDMA (WCDMA)

The worldwide introduction of WCDMA took place in 2001 and 2002, starting in Japan and continuing to Europe. In the U.S, several 3G alternatives will be available. GSM and TDMA operators can evolve toward EDGE, with WCDMA as a possible step, while cdmaOne operators can evolve toward cdma2000 systems.

WCDMA, as specified by the third-generation partnership project (3GPP), is a 3G system operat-

Figure 1. Comparison among the four generation communication system

1G	2G	3G	4G
Basic mobility	Advance mobility "roaming"	Seamless roaming	IP based mobility
Basic service	Various services "data exchange"	Service concept and model	Extremely high data rates
Incompatibility	Headed for global solution	Global solution	Perfect telecom, datacom convergence

ing in 5 MHz of bandwidth. Variable spreading and multicode operation is used to support a multitude of different radio access bearers. Different service classes are supported by an advanced quality-of-service (QoS) support. Data rates of up to 384 kbps for wide area coverage are provided (Bi, Zysman, & Menkes, 2001; Stallings, 2001; Dixit et al., 2001).

EDGE is an evolution of GPRS with data rates of up to 60 kbps per time-slot together with improved spectrum efficiency. EDGE uses higher-order modulation together with link adaptation and incremental redundancy to optimize the radio bearer to the radio connection characteristics. Currently, additions in the form of a new set of radio access bearers to align EDGE toward WCDMA are being standardized within the R5 of the 3GPP standards. The same service classes as in the WCDMA and the same interface to the core network will be used (Dornan, 2000; Ephremides et al., 2000; Fabri, Worrall, & Kondoz, 2000; Fasbender & Reichert, 1999; Flament et al., 1999; Frodigh, Parkvall, Roobol, Johansson, & Larsson, 2001).

CDMA 2000 (Evolution of IS-95 CDMA)

cdmaOne has evolved into cdma2000 and is available in two flavors, 1x and 3x. The former uses the same 1.25 Mhz bandwidth as cdmaOne and supports up to approximately 600 kbps, while the latter is a multi-carrier system using 3.75 Mhz and supporting approximately 2 Mbps at the moment, the focus on 3x is very limited. As a complement to 1x, the 3GPP2 has recently specified 1xEV-DO (1x Evolution-Data Only). 1xEV-DO uses a separate 1.25 Mhz carrier and supports best-effort data traffic only, using a new air interface compared to cdma2000 carrier. The peak rate in the 1x EV-DO downlink is almost 2.5 Mbps, excluding overhead. Phase two of the 1x evolution, known as 1xEV-DV (1x Evolution-Data and Voice), is currently being discussed within 3GPP2 and there are a number of proposals under consideration. The purpose is to specify an extension to cdma2000 1x in order to support high-rate data and voice on the same carrier (Bi, Zysman, & Menkes, 2001; Stallings, 2001; Dixit et al., 2001).

TD-SCDMA (Time Division-Synchronous CDMA)

UTRA TDD was developed to harmonize with the FDD component. This was achieved by harmonization of important parameters of the physical layer and a common set of protocols in the higher layers are specified for both FDD and TDD (Flament et al., 1999). TD-SCDMA has significant commonality with UTRA TDD. TD-SCDMA combines TDMA system with an adaptive CDMA component. TD-SCDMA eliminates the uplink/downlink interference, which affects other TDD methods by applying

"terminal synchronization" techniques, so TD-SCDMA can support of all radio network scenarios (Wide Area - Macro, Local Area - Micro, Hot Spots - Pico and Corporate Networks) to provide full service coverage. In this way, TD-SCDMA stands alongside W-CDMA and CDMA2000 as a fully-fledged 3G standard (Wireless Ethernet Compatibility Alliance).

DECT

DECT (digital enhanced cordless telecommunications) (European Telecommunications Standards) is a common standard for cordless personal telephony established by ETSI. DECT is used for those cordless communication systems which supports only indoor and pedestrian environment, but it does not allow full network coverage so is not satisfied to all requirements of third generation system.

UWC-136 (Evolution of IS-136)

DECT is based on TDMA. Different from UWC-136 (also based on TDMA), which uses two separate bandwidths (200 kHz provides medium bit rates up to 384 Kb/s and 1.6 MHz provides highest bit rates up to 2 Mb/s), DECT uses only one carrier with 1.728 MHz bandwidth. Variable bit rates are achieved by allocating different numbers of basic channels to a user. TDMA is flexible for TDD with asymmetric services and the training sequence is optimized for high bit rate services (Veerakachen, Pongsanguansin, & Sanguanpong, 1999).

In the fourth generation mobile communication (4G mobile), the combination and convergence of the different world's information technology industry, media industry and telecommunications will incorporate communication with information technology. As a result, mobile communications together with information technology will penetrate into the various fields of the society. In the future, 4G mobile (global and ubiquitous) communications will make people free from spatial and temporal constraints. Versatile communication systems will also be required to realize customized services based on diverse individual needs. The user outlook are increasing with regard to a large variety of services and applications with diverse degree of quality of service (QoS), which is related to delay, data rate, and bit error requirements. Therefore, seamless services and applications via different access systems and technologies that take full advantage of the use of available spectrum will be the driving forces for future developments.

Wireless Local Area Networks

Wireless local area networks (WLANs) based on the different versions of IEEE 802.11 standard have been around for some years in the 2.4 GHz ISM licensed band. Data rates up to 11 Mbps (802.11b) with reasonable indoor coverage have been offered, another licensed band is the microwave ovens which operates on 2.45 GH.

The ISM band is used in:

- 802.11
- Bluetooth
- Spread spectrum cordless phone

In an attempt to attain higher data rates new standards were proposed to operate on 5 GHz band with a rate up to 54 Mbps. Products based on two different standards, HIPERLAN 2 (H2) and IEEE 802.11a, The physical layers of the two are more or less identical with some differences where hiperlan2 uses several types of preambles but only PLCP is used in 802.11a to maintain synchronization and in modulation schemes (hiperlan2 supports 7 modes while 80.11a supports 8), with a carrier spacing of 20 MHz, OFDM modulation, and data rates up to 54 Mbps. The difference is the MAC protocol, where Hiperlan 2 has a more advanced protocol supporting QoS and mobility in a consistent way (Bi, Zysman, & Menkes, 2001).

802.11g is proposed to support high rates too up to 54 Mbps, Networks employing 802.11g operate at radio frequencies between 2.400 GHz and 2.4835 GHz, the same band as 802.11b (11 Mbps). But the 802.11g specification employs orthogonal frequency division multiplexing (OFDM), the modulation scheme used in 802.11a, to obtain higher data speed.

IEEE 802.11x

The current IEEE 802.11 (IEEE, 1999) is known to lack a viable security mechanism. In order to address this issue, IEEE has proposed the robust security network (RSN). RSN approved the 802.1x standard to provide strong authentication, access control, and key management. In a wireless environment, there are no physical perimeters, thus, the network access authentication must be provided. In this case, RSN provides mechanisms to restrict network connectivity to authorized entities only via 802.1x.

802.1x also provides various authentication methods such as one-time password and smartcards. It provides network access control for hybrid networking technologies (e.g., not only for wireless).

There are three entities specified in the IEEE 802.1x standard, including: supplicant, authenticator, and authentication server.

Wi-Fi

Wi-Fi (short for "wireless fidelity") is a term for certain types of wireless local area network (WLAN) that uses specifications in the 802.11 family (Vaughan-Nichols, 2003). The term Wi-Fi was created by an organization called the Wi-Fi Alliance, which oversees tests that certify product interoperability. A product that passes the alliance tests is given the label "Wi-Fi certified" (a registered trademark).

Originally, Wi-Fi certification was applicable only to products using the 802.11b standard (Ferro & Potorti, 2005). Today, Wi-Fi can apply to products that use any 802.11 standard. The 802.11 specifications are part of an evolving set of wireless network standards known as the 802.11 family. The particular specification under which a Wi-Fi network operates is called the "flavor" of the network. Wi-Fi has gained acceptance in many businesses, agencies, schools, and homes as an alternative to a wired LAN. Many airports, hotels, and fast-food facilities offer public access to Wi-Fi networks. These locations are known as hot spots. Many charge a daily or hourly rate for access, but some are free. An interconnected area of hot spots and network access points is known as a hot zone.

Unless adequately protected (Hole, Dyrnes, & Thorsheim, 2005), a Wi-Fi network can be susceptible to access by unauthorized users who use the access as a free Internet connection. The activity of locating and exploiting security-exposed wireless LANs is called war driving. An identifying iconography, called war chalking, has evolved. Any entity that has a wireless LAN should use security safeguards such as the wired equivalent privacy (WEP) encryption standard, the more recent Wi-Fi protected access (WPA), Internet protocol security (IPsec), or a virtual private network (VPN).

HiperLAN

It is a WLAN communication standard primarily used in European countries. There are two specifications: HiperLAN/1 and HiperLAN/2. Both have been adopted by the European Telecommunications Standards Institute (ETSI). The HiperLAN standards provide features and capabilities similar to 802.11. HiperLAN/1 provides communications at up to 20 Mbps in the 5-GHz range of the radio frequency (RF) spectrum. HiperLAN/ 2 is defined as a flexible Radio LAN standard designed to provide high speed access up to 54 Mbps to a variety of networks including 3G mobile core networks, ATM networks, and

IP-based networks, and also for private use as a wireless LAN system. Basic applications include data, voice and video, with specific Quality of Service (QoS) parameters taken into account. HiperLAN/2 systems can be deployed in offices, classrooms, homes, factories, hot spot areas like exhibition halls, and more generally where radio transmission is an efficient alternative or a complement to wired technology. It is worth noting that HiperLAN/2 has been developed in conjunction with the Japanese standards body, the Association of Radio Industries and Broadcasting.

HiperLAN/2 offers a number of advantages over 802.11a in that it incorporates quality of service (QoS) features. However, there is a long tradition of Europe developing standards which are not adopted because the US does something different. Most observers suggest that HiperLAN/2 will loose out to 802.11a, but that some of the features developed by ETSI will be incorporated in revised versions of the 802.11a standard.

WiMAX

WiMAX (Ghosh et al., 2005; Vaughan-Nichols, 2004; Hamalainen et al., 2002; Giuliano & Mazzenga, 2005) is a wireless industry coalition whose members organized to advance IEEE 802.16 standards for broadband wireless access (BWA) networks. WiMAX 802.16 technology is expected to enable multimedia applications with wireless connection and, with a range of up to 30 miles, enable networks to have a wireless last mile solution.

WiMAX was formed in April 2001, in preparation for the original 802.16 specification published in December of that year. According to the WiMAX forum, the group's aim is to promote and certify compatibility and interoperability of devices based on the 802.16 specification, and to develop such devices for the marketplace. Members of the organization include Airspan, Alvarion, Analog Devices, Aperto Networks, Ensemble Communications, Fujitsu, Intel, Nokia, OFDM Forum, Proxim, and Wi-LAN

WiMAX is a wireless industry coalition whose members organized to advance IEEE 802.16 standards for broadband wireless access (BWA) networks. WiMAX 802.16 technology is expected to enable multimedia applications with wireless connection and, with a range of up to 30 miles, enable networks to have a wireless last mile solution (Giuliano & Mazzenga, 2005; Ghavami, Michael, & Kohno, 2005). For reader reference, we are attaching a general specifications Table 1 for WiMAX standard.

Wireless Personal Area Networks

A WPAN (wireless personal area network) is a personal area network—a network for interconnecting devices centered around an individual person's work space—in which the connections are wireless. Typically, a wireless personal area network uses some technology that permits communication within about 10 meters—in other words, a very short range. One such technology is Bluetooth, which was used as the basis for a new standard, IEEE 802.15.

Table 1. WiMax (IEEE 802.16) specifications

Frequency range	Modulation	Multiple access	Duplex	Channel bandwidth	Number of channels	Peak data rate
2 GHz to 66GHz in various bands	BPSK, QPSK, 16QAM, 64QAM, QFDM, SC	TDMA/ OFDMA	TDD/ FDD	In accordance with local radio regulations	79	15 Mbit/s (5MHz channel) to 134 Mbit/s (28 MHz channel)

A WPAN could serve to interconnect all the ordinary computing and communicating devices that many people have on their desk or carry with them today—or it could serve a more specialized purpose such as allowing the surgeon and other team members to communicate during an operation.

A key concept in WPAN technology is known as plugging in. In the ideal scenario, when any two WPAN-equipped devices come into close proximity (within several meters of each other) or within a few kilometers of a central server, they can communicate as if connected by a cable. Another important feature is the ability of each device to lock out other devices selectively, preventing needless interference or unauthorized access to information.

The technology for WPANs is in its infancy and is undergoing rapid development. Proposed operating frequencies are around 2.4 GHz in digital modes. The objective is to facilitate seamless operation among home or business devices and systems. Every device in a WPAN will be able to plug in to any other device in the same WPAN, provided they are within physical range of one another. In addition, WPANs worldwide will be interconnected. Thus, for example, an archeologist on site in Greece might use a PDA to directly access databases at the University of Minnesota in Minneapolis, and to transmit findings to that database.

Infrared IR

802.11 also includes a specification for a physical layer based on infrared (IR) that is low enough that IR ports are standard on practically every laptop. IR is extremely tolerant of radio frequency (RF) interference because radio IR is unregulated. Product developers do not need to investigate and comply with directives from several regulatory organizations throughout the world and unauthorized users connecting to a network. Light can be confined to concerns. This comes at a price. IR LANs rely on scattering light off the ceiling, so range is much shorter. This discussion is academic, however.

No products have been created based on developed by the Infrared Data Association (IrDA), not 802.11. Even if products were created around the IR PHY, the big drivers to adopt 802.11 are flexibly penetrating solid objects. Using infrared light instead of radio waves seems to have several advantages. IR ports are less expensive than radio transceivers—in fact, the cost waves operate at a totally different frequency. This leads to a second advantage: Security concerns regarding 802.11 are largely based on the threat of conference room or office by simply closing the door. IR-based LANs can offer some of the advantages of flexibility and mobility but with less security the IR PHY. The infrared ports on laptops comply with a set of standards and mobility, which are better achieved by radio's longer range.

IrDA

IrDA (Infrared Data Association) is an industry-sponsored organization set up in 1993 to create international standards for the hardware and software used in infrared communication links (Williams, 2000; Vitsas & Boucouvalas, 2003). In this special form of radio transmission, a focused ray of light in the infrared frequency spectrum, measured in terahertz, or trillions of hertz (cycles per second), is modulated with information and sent from a transmitter to a receiver over a relatively short distance. Infrared radiation (IR) is the same technology used to control a TV set with a remote control.

IrDA has a set of protocols covering all layers of data transfer and, in addition, has some network management and interoperability designs. IrDA protocols have IrDA DATA as the vehicle for data delivery and IrDA CONTROL for sending the control information. In general, IrDA is used to provide wireless connectivity technologies for devices that would normally use cables for

connectivity (Robertson, Hansen, Sorensen, & Knutson, 2001).

IrDA is a point-to-point, narrow angle, ad-hoc data transmission standard designed to operate over a distance of 0 to 1 meter and at speeds of 9600 bps to 16 Mbps. Adapters now include the traditional upgrades to serial and parallel ports. In the IrDA-1.1 standard, the maximum data size that may be transmitted is 2048 bytes and the maximum transmission rate is 4 Mbps (Vitsas & Boucouvalas, 2002).

HomeRF

HomeRF (HomeRF) is a subset of the International Telecommunication Union (ITU) and primarily works on the development of a standard for inexpensive RF voice and data communication. The HomeRF Working Group has also developed the Shared Wireless Access Protocol (SWAP). SWAP is an industry specification that permits PCs, peripherals, cordless telephones and other devices to communicate voice and data without the usage of cables. It uses a dual protocol stack: DECT for voice, and 802.11 packets for data. It is robust, reliable, and minimizes the impact of radio interference. Its target applications are home networking, as well as remote control and automation.

Bluetooth

Bluetooth (Chiasserini, Marsan, Baralis, & Garza, 2003) is a high-speed, low-power microwave wireless link technology, designed to connect phones, laptops, PDAs and other portable equipment together with little or no work by the user. Unlike infrared, Bluetooth does not require line-of-sight positioning of connected units. The technology uses modifications of existing wireless LAN techniques but is most notable for its small size and low cost. Whenever any Bluetooth-enabled devices come within range of each other, they instantly transfer address information and establish small networks between each other, without the user being involved. To a large extent, Bluetooth have motivated the present WPAN attempts and conceptualizations; moreover, it constitutes the substance of the IEEE 802.15.1 WPAN standard. For reader reference, we are attaching general specifications in Table 2 for Bluetooth.

Bluetooth and 3G

Third-generation (3G) mobile telephony networks are also a familiar source of hype. They promise data rates of megabits per cell, as well as the "always on" connections that have proven to be quite valuable to DSL and cable modem customers. In spite of the hype and press from 3G equipment vendors, the rollout of commercial 3G services has been continually pushed back.

At the same time as 3G standards are being introduced, other air interfaces have been developed and standardized. First of all, Bluetooth is already available, enabling devices to communicate over short distances. The strength of Bluetooth is low power consumption and a design enabling low-cost implementations. Bluetooth was integrated into mobile phones, laptop computers, PDAs, and so forth. The first version of Bluetooth offers up to 700 kbps, but higher data rates, up to approxi-

Table 2. Bluetooth specifications

Frequency range	Modulation	Multiple access	Duplex	Channel bandwidth	Number of channels	Peak data rate
2402 MHz to 2480 MHz	GFSK	FHSS	TDD	1 MHz	79	723,2 kbit/s

mately 10 Mbps, are currently being standardized for later releases.

In contrast to Bluetooth and 3G, equipment based on the IEEE 802.11 standard has been an astounding success. While Bluetooth and 3G may be successful in the future, 802.11 is a success now. Apple initiated the pricing moves that caused the market for 802.11 equipment to explode in 1999. Price erosion made the equipment affordable and started the growth that continues today. IEEE

802.15 WPAN

802.15 is a communications specification that was approved in early 2002 by the Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA) for wireless personal area networks (WPANs) (Chiasserini et al., 2003). The initial version, 802.15.1, was adapted from the Bluetooth specification and is fully compatible with Bluetooth 1.1.

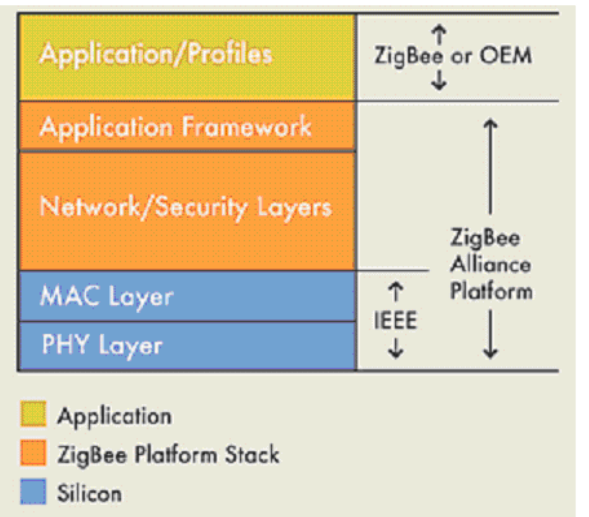
The IEEE 802.15 Working Group proposes two general categories of 802.15, called TG4 (low rate) and TG3 (high rate). The TG4 version provides data speeds of 20 Kbps or 250 Kbps. The TG3 version supports data speeds ranging from 11 Mbps to 55 Mbps. Additional features include the use of up to 254 network devices, dynamic device addressing, support for devices in which latency is critical, full handshaking, security provisions, and power management. There will be 16 channels in the 2.4-GHz band, 10 channels in the 915-MHz band, and one channel in the 868-MHz band (HomeRF; Shiraishi, 1999).

The IEEE plans to refine the 802.15 specification to work with the specification and description language (SDL), particularly SDL-88, SDL-92, and SDL-2000 updates of the International Telecommunication Union (ITU) recommendation Z.100.

ZigBee

Wireless networks present colossal opportunities; at the same time, they are constrained by substantial development challenges. ZigBee™ provides standard-based solution for lightweight wireless networks based on the IEEE 802.15.4 (You, Park, Ju, Kwon, & Cho, 2001). Figure 2 provides clear elaboration for the ZigBee stack

Figure 2. ZigBee standard stack definition



definition. While IEEE has defined the standards for MAC and PHY layers, the ZigBee™Alliance has added standards for network, security, and application layers.

Our readers may also refer to the general description Table 3 of ZigBee.

To conclude our section, a comparison in Table 4 among a set of frequently used personal wireless not working up to the standard is provided. This table states the standard bandwidth, power consumption, protocol stack size, stronghold, and finally draws preferred applications for each standard.

APPLICATIONS AND SERVICES

Overview

The concept of mobile multimedia services was first introduced in 1992 when the ITU realized that mobile communications were playing an increasingly important role. It began working on a project called *FPLMTS* (Future Public Land Mobile Telecommunications System) aiming to unite the world under a single standard.Given the fact, however, that this acronym is difficult to pronounce, it was subsequently renamed *International Mobile Telecommunications – 2000* (IMT-2000).

Table 3. ZigBee (IEEE 802.15.4) specifications

Frequency range	Modulation	Multiple access	Duplex	Channel bandwidth	Number of channels	Peak data rate
2.4 GHz to 2.4835 GHz (World) 902 MHz to 928 MHz (America) 868.3 MHz (Europe)	BPSK (868/915 MHz), OQPSK (2.4 GHz)	CSMA/ CA	TDD	5 MHz	1 (868 MHz) 10 (915 MHz) 16 (2.4 GHz)	20 kbit/s (868 MHz) 40 kbit/s (915 MHz) 250 kbit/s (2.4 GHz)

Table 4. Wireless technology comparison table

Standard	Bandwidth	Power Consumption	Protocol Stack Size	Stronghold	Applications
Wi-Fi	Up to 54Mbps	400+mA TX, standby 20mA	100+KB	High data rate	Internet browsing, PC networking, file transfers
Bluetooth	1Mbps	40mA TX, standby 0.2mA	~100+KB	Interoperability, cable replacement	Wireless USB, handset, headset
ZigBee	250kbps	30mA TX, standby 1uA	4”32KB	Long battery life, low cost	Remote control, battery-operated products, sensors

IMT-2000 is a single family of compatible standards defined by a set of ITU-R Recommendations. The main objectives for IMT-2000 are (UMTS Forum Report, 2000):

- High data rates, 144 Kbps/384 Kbps for high mobility users with full coverage and 2 Mbps for low mobility users with limited coverage,
- Capability for multimedia application and all mobile applications,
- High spectrum efficiency compared to existing systems,
- High flexibility to introduce new services,
- High degree of commonality of design worldwide, and
- Use of a small pocket terminal with seamless global roaming.

The main applications will not be on only traditional voice communications, but also on services such as e-mail, short messages, multimedia, simultaneous voice and data and the broadband integrated service digital network (B-ISDN) access (UMTS Forum Report, 2000). Although the PC has been the dominant Internet client, soon mobile phones and personal digital assistants (PDAs) will outnumber PCs and be the major source of Internet connections. With the third-generation (3G) network, mobile device users won't have to worry about bandwidth. People will access video on their phones as easily as they do on their PCs. In fact, the development of GSM Networks and terminals to support more advanced data bearer technologies has allowed the introduction of new exciting data services. These technologies allow a greater bandwidth and more capable execution environment permitting the development of mobile applications. The world has become increasingly computer centric, and computer applications are now used for a number of tasks such as communications, financial management, information retrieval, entertainment, and game playing.

It is a natural progression for the user to expect these applications to be available for them on their mobile terminal. The initial developments in mobile applications were basic, running on the GSM SIM Card using SIM toolkit interfacing through capable terminals and using SMS to communicate with the application infrastructure. These were followed by the introduction of browsers utilizing special mobile protocols (WAP, HDML, etc.) and the basic data capabilities of the GSM terminals of the time. These allow the user to use their mobile terminal to access basic text format content such as news, sport, entertainment, and information, among others.

The introduction of high bandwidth capability allows for richer applications, and the packet switched nature of GPRS networks allows for more efficient applications. They will smooth the progress of the introduction of true multimedia services such as multimedia messaging service (MMS) which will allow the user to send and receive messaging containing pictures, images, sound, and text. New network features, such as location servers, can allow the mobile applications to be improved to tailor the way they work, and improve the value to the user. Mobile commerce is the effective delivery of electronic commerce into the consumer's hand, anywhere, using wireless technology. This advance has the power to transform the mobile phone into a "mobile wallet." Already, major companies have begun to establish partnerships with banks, ticket agencies, and top brands to take benefit of the retail outlet in the consumer's hand.

Location based services provide personalized services to the subscriber based on their current position. Two categories of methods can be used to find the location of the subscriber: basic and advanced. Basic positioning methods are based on the information of the cell the subscriber is using (cell ID). This can be used alone, or together with other information available in the network such as timing advance and network measurement reports. This information is available for

all handsets. Advanced techniques will be available in new handsets such as enhanced observed time difference and assisted GPS which uses the GSM network to help the reception of the freely available GPS system. The division of position technologies above is based on accuracy of the positioning method. Other factors also very important are for example complexity of the system, availability of the positioning technology in the network (e.g., assisted-GPS may not be available everywhere) and the investment needed on the network side and in handsets. Mobile video streaming is another application of whole big set of applications in field of mobile multimedia. The facts of high bit rate and smart phone availability will put a hundred of applications ongoing for mobile multimedia.

Last but not least, interesting new technology, the so-called mobile TV, has strongly emerged in the mobile multimedia field as taking breath and killer application. Interest in television services over cellular or broadband wireless networks is intensifying as operators seek a new and high margin application. Philips Semiconductors is predicting that, within a decade, the majority of its television chips will go into cell phones, not conventional television sets.

Classification of Services

Mobile multimedia services aim to combine the Internet, telephones, and broadcast media into a single device (UMTS Forum Report, 2000). To achieve this, IMT-2000 systems have been designed with six broad classes of service in mind. None of them are yet set in hardware but they are useful for regulators planning coverage and capacity, and perhaps for people buying terminals when they finally become available.

It's likely that 3G devices will be rated according to the types of service they can access, from a simple phone to a powerful computer. Three of the service classes are already present to some extent on 2G networks, while three more are new and involve mobile multimedia. In order of increasing data rate:

- **Voice:** Even in the age of high-speed data, this is still regarded as the "killer app" for the mobile market. 3G will offer call quality at least as good as the fixed telephone network, possibly with higher quality available at extra cost. Voicemail will also be standard and eventually integrated fully with email through computerized voice recognition and synthesis.
- **Messaging:** This is an extension of paging, combined with Internet e-mail. Unlike the text-only messaging services built into some 2G systems, 3G will allow email attachments. It can also be used for payment and electronic ticketing.
- **Switched data:** This includes faxing and dial-up access to corporate networks or the Internet. With always-on connections available, dial-up access ought to be obsolete, so this is mainly included to support legacy equipment. In 3G terms, legacy means any product that doesn't support a fully packet-switched network.
- **Medium multimedia:** This is likely to be the most popular 3G service. Its downstream data rate is ideal for Web surfing. Other applications include collaborative working, games, and location-based maps.
- **High multimedia:** This can be used for very high-speed Internet access, as well as for high-definition video and CD-quality audio on demand. Another possible application is online shopping for "intangible" products that can be delivered over the air; for example, music or a program for a mobile computer.
- **Interactive high multimedia:** This can be used for fairly high-quality videoconferencing or videophones, and for telepresence, a combination of videoconference and collaborative working.

The data rates of these services are shown in Table 1, together with their level of asymmetry and switching mode. These services refer to three basic levels by which these services are structured according to the dependencies among these services:

1. **Basic level services:** Those services form the building blocks of other more complex applications and services, such as voice messaging, data retrieval, video, and so forth, and can be used as stand-alone services or form the ingredients of higher level services and applications.
2. **Value added services:** Those services form the intermediate level services formed by one or more basic level services. VAS offer optimized functionality to suit the needs of diverse professional groups. Examples of such services are wireless home networking, high data rate PAN, high density networks, P2P communication collaboration, Internet/ Intranet access, video conferencing, telemetry, location based services, payments, and UMS.
3. **High level applications:** Those address the specific requirements and interests of professional or consumer user groups. These are functionally stand-alone and serve the full range of user needs supporting services forms them. Examples can be business applications, transaction management, information applications, entertainment applications, telematics, construction, electronic healthcare, provision, e-government, e-learning, wireless home networking, and so on.

The above taxonomy refers to the functionality of the service and the group of users it targets. In the context of mobile multimedia, the basic level services refer to the collection, sharing and exchange of multimedia data, while the value added level refers to the provision and distribution of the multimedia information and the high level applications is concerned about the usage and consumption of the data.

THE FUTURE OF MOBILE MULTIMEDIA

Future generation mobile terminals will start to incorporate ubiquitous network functionality by efficiently dealing with a massive amount of communication modes and various multimedia applications. Further, these terminals will also need adaptive behavior to intelligently manage the computational resources that will be distributed and shared across the environing systems. Researchers should find integrated research platform aiming to resolve the fundamental technological issues for future mobile terminals allowing the true ubiquitous network environment to become a reality. Complexity, cost, power consumption, high throughput at low latency, and flexibility are the five primary hurdles in developing a mobile terminal.

In addition, many types of objects as well as people will have network functions and communicate with each other through networks. Therefore, different communication relationships such as person to person, machine to machine and mainly machine to person and vice versa, will determine mobile and wireless communications in the future.

Given the increasing demand for flexibility and individuality in society, the mean for the end-user might be assessed. Potentially, the value would be in the diversity of mobile applications, hidden from the complexity of the underlying communications schemes. This complexity would be absorbed into an intelligent personality management mechanism, which would learn and understand the needs of the user, and control the behavior of their reconfigurable and open wireless terminals accordingly in terms of application behavior and access to future support services.

In the future wireless service provision will be characterized by global mobile access (terminal and personal mobility), high quality of services (full coverage, intelligible, no drop and no/lower call blocking and latency), and easy and simple access to multimedia services for voice, data, messages, video, WWW, GPS, and so forth, via one user terminal.

End-to-end secured services will be fully coordinated via access control, authentic use of biometric sensors and/or smart card and mutual authentication, data integrity, and encryption. User added encryption feature for higher level of security will be part of the system.

Considering how second-generation systems have evolved by adding more and more system capabilities and enhancements to make them resemble the capabilities of 3G systems; it is possible that with third-generation systems there may be a continuum of enhancements that will render those systems practically indistinguishable from future generation systems. Indeed, it is expected that it will be more difficult to identify distinct generation gaps, and such a distinction may only be possible by looking back at some point in the future.

Progress has also been made in the development of other signal processing techniques and concepts for use in tomorrow's wireless systems. These include smart antennas and diversity techniques, better receivers, and hand over and power control algorithms with higher performance (Bi, Zysman, & Menkes, 2001).

ACKNOWLEDGMENTS

The authors of this chapter would like to express their heartfelt thanks to Huang Hsin-Yi and Lin Yu-Hua (Monica) for their proofreading our chapter.

REFERENCES

Bi, Q., Zysman, G.I., & Menkes, H. (2001). Wireless mobile communications at the start of the 21st century. *IEEE Communications Magazine,* 110-116.

Blair, G.S., Coulson, G., Davies, N., Robin, P., & Fitzpatrick, T. (1997). Adaptive middleware for mobile multimedia applications. In *Proceedings of the 7th International Conference on Network and Operating System Support for Digital Audio and Video (Nossdav'97),* St Louis, Missouri (pp. 259-273)

Brown, M.G., & Syfrig, H. (1996). Follow-me-video in a distributed computing environment. In the *Proceedings of the 3rd International Workshop on Mobile Multimedia Communications.* Princeton, NJ: Plenum Publishers.

Bruno, R., Conti, M., & Gregori, E. (2001). WLAN technologies for mobile ad hoc networks. In *Proceedings of the 34th Hawaii International Conference on System Sciences.*

Bull, D., Canagarajah, N., & Nix, A. (1999). *EEE Communications Magazine, 39*(2).

Chiasserini, C.F., Marsan, M.A., Baralis, E., & Garza, P. (2003). Towards feasible topology formation algorithms for Bluetooth-based WPAN's. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences.*

Collins, D., & Smith, C. (2001). *3G wireless networks.* McGraw-Hill Professional.

Dixit, S., Guo, Y., & Antoniou, Z. (2001). Resource management and quality of service in third generation wireless network. *IEEE Communication Magazine, 39*(2).

Dornan, A. (2000). *The essential guide to wireless communications applications: From cellular systems to WAP and m-commerce.* Prentice Hall PTR.

Ephremides, A., et al. (2000). *Wireless technologies and information networks.* International Technology Research Institute, World Technology (WTEC) Division, WTEC Panel Report.

Fabri, S.N., Worrall, S.T., & Kondoz, A.M. (2000). Video communications over mobile networks. *Communicate 2000*, Online Conference, London.

Fasbender, A., & Reichert, F. (1999). Any network, any terminal, anywhere. *IEEE Personal Communications*, 22-29.

Ferro, E., & Potorti, F. (2005). Bluetooth and Wi-Fi wireless protocols: A survey and a comparison. *Wireless Communications*, *12*(1), 12-26.

Flament, M., et al. (1999). An approach to 4th generation wireless infrastructures: Scenarios and key research issues. *IEEE VTC 99*, Houston, TX.

Frodigh, M., Parkvall, S., Roobol, C., Johansson, P., & Larsson, P. (2001). Future generation wireless networks, *IEEE Personal Communications.*

Ghavami, M., Michael, L.B., & Kohno, R. (2005). Ultra wideband signals and systems in communications engineering, *Electronics Letters*, *41*(25).

Ghosh, A., et al. (2005). Broadband wireless access with WiMax=8O2.16: Current performance benchmarks and future potential. *IEEE Commun. Mag.*, *43*(2), 129-136.

Giuliano, R., & Mazzenga, F. (2005). On the coexistence of power-controlled ultrawide-band systems with UMTS, GPS, DCS 1800, and fixed wireless systems, *IEEE Trans. Veh. Technol.*, *54*(1), 62-81.

Hamalainen, M., et al. (2002). On the UWB system coexistence with GSM900, UMTS=WCDMA, and GPS, *IEEE J. Sel. Areas Commun.*, *20*(9), 1712-1721.

Hole, K.J., Dyrnes, E., & Thorsheim, P. (2005). Securing Wi-Fi networks. *Computer, 38*(7), 28-34.

IEEE 802.11 (1999). *Local and metropolitan area networks: Wireless LAN medium access control (MAC) and physical specifications.* ISO/IEC 8802-11:1999

Robertson, M.G., Hansen, S.V., Sorenson, F.E., & Knutson, C.D. (2001). Modeling IrDA performance: The effect of IrLAP negotiation parameters on throughput. *Proceedings of the 10th International Conference on Computer Communications and Networks* (pp. 122-127).

Shiraishi, Y. (1999). Communication network: Now and in future. *OKI Technical Review*, Issue 162, *65*(3).

Stallings, W. (2001). *Wireless communications and networks.* Prentice Hall.

UMTS Forum Report No. 10 (2000). *Shaping the mobile multimedia future.*

UMTS Forum Report No. 11 (2000). *Enabling UMTS Third Generation Services and Applications.*

UMTS Forum Report No. 3 (1998). *The impact of license cost levels on the UMTS business case.*

UMTS Forum Report No. 4 (1998). *Considerations of licensing conditions for UMTS network operations.*

UMTS Forum Report No. 9 (2000). *The UMTS third generation market: Structuring the service revenue opportunities.*

Vaughan-Nichols, S.J. (2003). The challenge of Wi-Fi roaming, *Computer, 36*(7), 17-19.

Vaughan-Nichols, S.J. (2004). Achieving wireless broadband with Wi-Max. *IEEE Computer, 37*(6), 10-13.

Veerakachen, W., Pongsanguansin, P., & Sanguanpong, K. (1999). Air interface schemes for IMT-2000. *NECTEC Technical Journal, 1*(1).

Vitsas, V., & Boucouvalas, A.C. (2002). IrDA IrLAP protocol performance and optimum link layer parameters for maximum throughput, *Global Telecommunications Conference, 2002. GLOBECOM'02, 3*, 2270-2275.

Vitsas, V., & Boucouvalas, A.C. (2003). Optimization of IrDA IrLAP link access protocol. *IEEE Transactions on Wireless Communications, 2*(5), 926-938.

Williams, S. (2000). IrDA: Past, present and future. *Personal Communications, 7*(1), 11-19.

You, Y.-H., Park, C.-H., Ju, M.-C., Kwon, K.-W. & Cho, J.-W. (2001). Adaptive frequency hopping scheme for interference-limited WPAN applications. *Electronics Letters, 37*(15), 976-978.

This work was previously published in Business Data Communications and Networking: A Research Perspective, edited by J. Gutiérrez, pp. 128-150, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.7
Interactive Multimedia File Sharing Using Bluetooth

Danilo Freire de Souza Santos
Federal University of Campina Grande, Brazil

José Luís do Nascimento
Federal University of Campina Grande, Brazil

Hyggo Almeida
Federal University of Campina Grande, Brazil

Angelo Perkusich
Federal University of Campina Grande, Brazil

INTRODUCTION

In the past few years, industry has introduced cellular phones with increasing processing capabilities and powerful wireless communication technologies. These wireless technologies provide the user with mechanisms to easily access services, enabling file sharing among devices with the same technology interfaces (Mallick, 2003). In the context of electronic commerce, which demands new techniques and technologies to attract consumers, these wireless technologies aim to simplify the shopping process and provide up-to-date information about available products.

In order to exemplify the application of mobile and wireless technologies to satisfy these new commerce functionalities and needs, we present in this article the interactive multimedia system (IMS). IMS is a system for sharing multimedia files between servers running on PCs, and client applications running on mobile devices. The system was conceived initially to be deployed in CDs/DVDs and rental stores to make available product information in a simple and interactive way.

In a general way, the system allows a user to obtain information about available products through a mobile device. Then, a user can listen or watch parts (stretches) of available videos or songs. For that, the user needs to enter the store, choose a product in the store shelf, and type its identity code in the mobile device, choosing which

music (or video) to listen to (or watch).

The IMS system has a client/server architecture, where the server was developed in C++ for the Windows operating system and the client application was developed in C++ for the Symbian operating system, which is a mobile device operating system mainly used in smart phones. Client/server communication is performed based on Bluetooth wireless technology. Bluetooth is suitable for this kind of application because it has a satisfactory transmission rate with enough range, and it is also supported by more than 500 million mobile devices (Bluetooth Official Website, 2006).

The rest of this article is organized as follows. In next section we present a background of the main technologies used in this project. We then present the architecture of the proposed system and describe how the system works, before discussing future trends of mobile multimedia systems and offering final remarks.

BACKGROUND

This section provides an overview of the main technologies used in the IMS development. More specifically, we outline the Bluetooth wireless technology and the programming language used with the Symbian Operational System.

Bluetooth

To provide communication between devices, the IMS client/server architecture uses the Bluetooth wireless technology. Bluetooth is a short-range wireless technology present in a large number of smart phones of the Symbian OS Series 60 platform. It is suitable for fast file exchange, including text files, photo files, and short video files. Bluetooth technology covers a distance of about 10 meters for class 2 devices (most common devices), and each server (or master) can be connected to up to seven slaves in its coverage area (Mallick, 2003). Another important feature of Bluetooth is its lower power consumption, around 2.5 mW at most, which reinforces its use in embedded devices.

With Bluetooth, it is possible to use two kinds of connections: ACL (asynchronous connectionless) and SCO (synchronous connection oriented) (Andersson, 2001). ACL links are defined for data transmission, supporting symmetrical and asymmetrical packet-switched connections. In this mode, the maximum data rate could be 723 kbps in one direction and 57.6 kbps in the other direction, and these rates are controlled by the master of a cell. SCO links support only a symmetrical, circuit-switched, point-to-point connection for primarily voice traffic. The data rate for SCO links is limited to 64 kbps, and the number of devices connected at the same time with the master is restricted to three devices.

In the Bluetooth protocol stack, some profiles that implement some kind of particular communication partner are defined. The general profiles in Bluetooth stack are: GAP (generic access profile), SDAP (service discovery application profile), SPP (serial port profile), and GOEXP (generic object exchange profile) (Forum Nokia, 2003). GAP defines generic procedures related to discovery of Bluetooth devices and links management aspects of connecting Bluetooth devices. SDAP defines features and procedures to allow an application in a Bluetooth device to discover services of another Bluetooth device. With SPP used in ACL links, it is possible to emulate serial cable connections using RFCOMM (RS232 Serial Cable Emulation Profile) between two peer devices. RFCOMM emulates RS-232 (Serial Cable Interface Specification) signals and can thus be used in applications that are formerly implemented with a serial cable. The GOEXP profile defines protocols and procedures that should be used by applications requiring object exchange capabilities.

Symbian

Symbian is an operating system specifically designed for mobile devices with limited resources, such as memory and processor performance. The programming language C++ for Symbian provides a specific API (application program interface), with new features for the programmer that allow access to services such as telephony and messaging (Stichbury, 2004). Also, the Symbian C++ API enables programmers to efficient deal with multitasking and memory functions. These functions reduce memory-intensive operations. Symbian OS is event driven rather than multi-thread. Although multi-thread operations are possible, they potentially create kilobytes of overhead per thread. Services in Symbian OS are provided by servers through client/server architectures. For developing applications, Symbian offers an application framework, which constitutes a set of core classes that are the basis and structure of all applications.

The Symbian OS architecture can be described by a layered approach, as illustrated in Figure 1.

The layers can be defined as follows:

- **User Interface (UI):** Can be specifically defined per vendor or per family of mobile devices, such as Series 60 platform devices;
- **Application Module:** Allows access to applications' built-in functionality concerned with data processing and not how it is presented for the user;
- **System Module:** Contains the set of OS APIs; and
- **Kernel:** The core of the operational system and cannot be directly accessed by user programs.

The mobile application was implemented using a Series 60 platform (Series 60 Website, 2006; Edwards & Barker, 2004). Series 60 is a complete smart phone-based UI design reference. It completes the Symbian OS architecture with a configurable graphical user interface library and a suite of applications, besides other general-purpose engines.

SYSTEM ARCHITECTURE

This section presents the IMS architecture. As introduced earlier, the IMS system has a client/server architecture, where the server was developed in C++ for the Microsoft Windows OS and the client application was developed in C++ for Symbian OS. Client/server communication is performed based on Bluetooth wireless technology, which can provide connection to up to seven users at the same time. Figure 2 illustrates a general system view.

According to this general view, the system can be divided into three specific modules: mo-

Figure 1. Symbian OS architecture

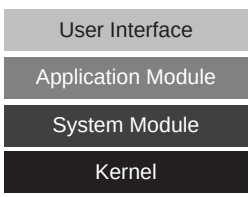
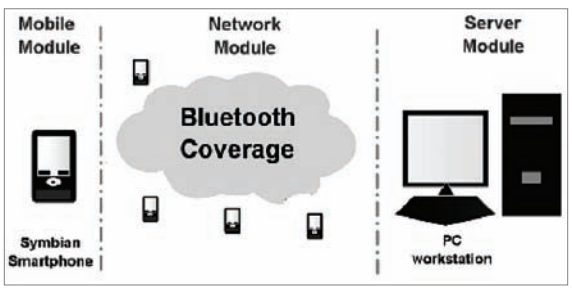


Figure 2. IMS general view



bile, network, and server. The mobile module is composed of the software running on mobile devices, such as smart phones. It offers a friendly and intuitive user interface, which is responsible for showing relevant information about available products. This information is obtained by typing the product code into the device. Then, the product description is returned with an option to run the multimedia file related to the product. Also, the mobile module controls a multimedia player installed together with the IMS software. Screenshots of the IMS mobile application are presented in Figure 3.

The network module controls the mobile network, in this case the Bluetooth wireless technology. To handle the connection, the network module offers an application layer protocol to manage the exchange of messages between the other two modules. This protocol is text based. Also, this module is responsible to control the transfer of files using the Bluetooth Serial Port Profile (SPP). The SPP profile is supported by Symbian (and Series 60 Platform) Bluetooth-enabled devices (Jipping, 2003). The system can support more than seven users by sharing the available Bluetooth communication ports at server side. This way, after using one communication port with one user, the network module switches the port to another user if necessary.

Finally, the server module is responsible for database and system management. Files and relevant information (description of video or song, time, author, etc.) are stored into a database. When required, the server sends relevant information about a specific product. After confirmation from the mobile software, the server sends the multimedia file.

SYSTEM EXECUTION

In this section we describe how IMS works. For this execution scenario, consider that the client has entered the IMS Bluetooth coverage area. Then, he/she has chosen a product to retrieve information (such as a DVD or CD) available at the store shelf. After that, as illustrated in Figure 4, the following steps are performed:

1. The user types the product identification code into the mobile device software. Then, the software running on the mobile device searches for the IMS server using Bluetooth technology. After finding the IMS server, the IMS network module establishes the connection between the mobile device and the server. Using a specific application protocol, the IMS software asks for information about that

Figure 3. IMS screenshots

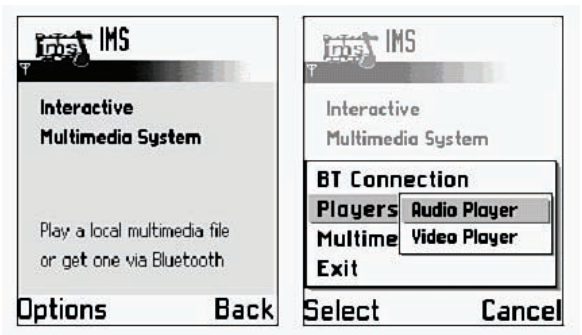
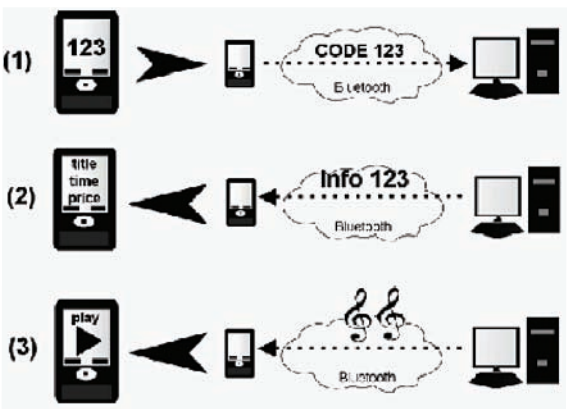


Figure 4. IMS operation



product code, such as audio/video stretches, release notes, and so forth.
2. If the information is available, the server module retrieves it from the database and returns it to the mobile device through the Bluetooth link. Then, the IMS software receives that information and displays it to the client, who will have available the description and the option to run a multimedia file.
3. When the client chooses to run the file, it is downloaded through the Bluetooth link and executed using a multimedia player managed by the IMS software.

FUTURE TRENDS

Today, the mobile multimedia area tends to investigate and develop mechanisms for carrying multimedia streams over wireless links, as described by Chen, Kapoor, Lee, Sanadidi, and Gerla (2004). Mobile multimedia systems are focused on providing multimedia streams directly to the mobile device in an efficient way. Future interactive multimedia systems will offer real-time multimedia streams, with more general services such as IPTV (television over IP) (Santos, Souto, Almeida, and Perkusich, 2006), radio, and so forth. Also, with the advent of smart phones with different wireless interfaces such as Wi-Fi and Bluetooth, those systems will support heterogeneous technologies offering different kinds of services for different kinds of devices.

Another prominent future trend is related to pervasive computing (Weiser, 1991; Saha & Mukherjee, 2003; Satyanarayanan, 2001). Within this context, users with mobile and wireless technology access can define their personal preferences to adapt systems and environments according to it. IMS, in the future, could be used in a pervasive infrastructure enabling file sharing according to personal preferences of users. IMS can be primarily used to delivery file in the pervasive environment, and in the future can be associated with variable bit rate multimedia streaming according to battery life of a mobile device or any kind of strategy based on the user's profile.

CONCLUSION

Mobile multimedia systems are becoming reality, and more and more accessible to ordinary users. These users, as potential consumers, are attracting industry and commerce attention, motivating research and development of several solutions in this area. Within this context, one of the relevant efforts is to promote multimedia file sharing.

To illustrate the large application of multimedia file sharing to support consumer activities, we presented in this article an interactive multimedia system. IMS provides attractive and interactive ways for users to obtain product information, which can be used in several commerce domains.

Interactive multimedia systems, such as IMS, are very useful and have great relevance for multimedia commerce. They could substitute traditional multimedia players installed in stores, offering a new way of interacting with clients using mobile devices.

REFERENCES

Andersson, C. (2001). *GPRS and 3G wireless applications.* New York: John Wiley & Sons.

Bluetooth Official Website. (2006). *Bluetooth technology benefits.* Retrieved April 28, 2006, from http://www.bluetooth.com/Bluetooth/Learn/Benefits/

Chen, L., Kapoor, R., Lee, K., Sanadidi, M.Y., & Gerla, M. (2004). Audio streaming over Bluetooth: An adaptive ARQ timeout approach. *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW'04).*

Edwards, L., & Barker, R. (2004). *Developing Series 60 applications—a guide for Symbian OS C++ developers.* Nokia Mobile Developer Series. Boston: Addison-Wesley.

Forum Nokia. (2003, April 4). *Bluetooth technology overview.* Retrieved March 1, 2005, from http://www.forum.nokia.com

Jipping, M. (2003). *Symbian OS communications programming.* New York: John Wiley & Sons.

Mallick, M. (2003). *Mobile and wireless design essentials.* New York: John Wiley & Sons.

Saha, D., & Mukherjee, A. (2003, March). Pervasive computing: A paradigm for the 21st century. *IEEE Computer,* 25-31.

Santos, D.F.S., Souto, S.F., Almeida, H., & Perkusich, A. (2006, April). An IPTV architecture using free software (in Portuguese). *Proceedings of the Brazilian Computer Society's Free Software Workshop in the International Free Software Forum,* Porto Alegre, Brazil.

Satyanarayanan, M. (2001). Pervasive computing: Vision and challenges. *IEEE Personal Communications, 8*(4).

Series 60 Web Site. (2006). *S60—about Series 60.* Retrieved April 28, 2006, from http://www.s60.com/about

Stichbury, J. (2004). *Symbian OS explained—effective C++ programming for smart phones.* New York: John Wiley & Sons.

Weiser, M. (1991). The computer for the 21st century. *Scientific American, 265*(3), 94-104.

KEY TERMS

ACL: Asynchronous connectionless.

API: Application program interface.

Bluetooth Link: Connection between two Bluetooth peers.

GAP: Generic access profile.

GOEXP: Generic object exchange profile.

IMS: Interactive multimedia system.

OS: Operational system.

SCO: Synchronous connection oriented.

SDAP: Service discovery application profile.

Smart Phone: Cell phone with special computer-enabled features.

SPP: Serial port profile.

UI: User interface.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 341-344, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.8
Security of Mobile Devices for Multimedia Applications

Göran Pulkkis
Arcada University of Applied Sciences, Finland

Kaj J. Grahn
Arcada University of Applied Sciences, Finland

Jonny Karlsson
Arcada University of Applied Sciences, Finland

Nhat Dai Tran
Arcada University of Applied Sciences, Finland

INTRODUCTION

Users of the Internet have become increasingly more mobile. At the same time, mobile users want to access Internet wireless services demanding the same quality as over a wire. Emerging new protocols and standards, and the availability of WLANs, cellular data and satellite systems are making the convergence of wired and wireless Internet possible. Lack of standards is however still the biggest obstacle to further development. Mobile devices are generally more resource constrained due to size, power, and memory. The portability making these devices attractive greatly increases the risk of exposing data or allowing network penetration.

Multimedia applications in mobile devices require support for continuous-media data types, high network and memory bandwidth, low power consumption, low weight and small size, and QoS (quality of service). Also security features like authentication and authorization for multimedia content as well as secure connectivity to multimedia sources with possibilities to verify the integrity and guarantee the confidentiality of delivered multimedia content are required (Havinga, 2000). A mobile user must also be able to roam between different networks, also between different types of networks (WLAN, cellular, etc.), and still maintain an ongoing secure multimedia application session.

A standard for a mobile multimedia system architecture has also been proposed (MITA, 2002). Two technologies, MBMS (Multimedia Broadcast/Multimedia Service [see MBMS, 2004]), and DVB-H (digital video broadcasting handheld [see ETSI EN 302 304, 2004]), are being developed for delivery of multimedia content to mobile devices.

To fulfill all security requirements for multimedia applications in a mobile environment, while still maintaining QoS, is a challenging issue. In this chapter, the security requirements and proposed solutions for fulfilling these requirements are discussed. Attention is paid to ongoing related research. However, in order to achieve a secure mobile multimedia environment, also basic mobile device security issues must be seriously taken into account.

BACKGROUND

A mobile computer is a computing device intended to maintain its functionality while moving from one location to another. Different types of mobile computers are:

- Laptops
- Sub-notebooks
- PDAs (personal digital assistants)
- Smartphones

These devices can be divided in two groups: handheld devices and portable PCs. Handheld devices, such as PDAs and smartphones, are pocket-sized computing devices with smaller computing, memory, and display capacity compared to basic desktop computers. Portable PCs such as laptops and sub-notebooks, however, don't significantly differ from the desktops on this area.

Mobile computing and mobility are generic terms for describing the ability to use mobile devices for connecting to and using centrally located applications and/or services over a wireless network. Mobile multimedia can be defined as a set of protocols and standards for exchanging multimedia information over wireless networks. Mobile multimedia user services are usually divided in three categories (MBMS, 2004):

- streaming services like real-time video and audio streams, TV and radio programs, and on-demand video services;
- file download services; and
- carousel services, for example, news delivery with timed updating of text, image, and video objects.

For setup of sessions using multimedia streaming services in IP-based networks is used a signaling protocol, the Session Initiation Protocol (SIP), which is an Internet standard adopted by the Internet Engineering Task Force (IETF). SIP is a text-based client server protocol, in which servers respond to SIP requests sent by clients. SIP entity types are user agents (SIP clients), proxy servers, redirect servers, and registrar servers. Two or more participants can establish a session consisting of multiple media streams. SIP also provides application level mobility, which includes personal, terminal, session, and service mobility. *Personal mobility* means that a user ID provides global accessibility. *Terminal mobility* means that a mobile end-user device can maintain streaming media sessions while moving within and between IP subnets. *Session mobility* means users can change terminals during streaming media sessions and still maintain their sessions. *Service mobility* means that users maintain their streaming media session while moving, changing terminal devices, and changing network service providers (Rosenberg et al., 2002).

SIP does not define any protocol for media transport. However, streaming services typically use the Real-time Transport Protocol (RTP) over UDP. RTP defines a standardized packet format for delivering audio and video over the Internet (Schulzrinne, Casner, Frederick, & Jacobson, 2003).

Multimedia applications for handheld devices require support for continuous-media data types, high network and memory bandwidth, low power consumption, low processing power, low memory capacity, low weight and small size, and quality of service. Therefore, for multimedia applications, most of the research has currently been focused on smoothness of the multimedia data delivery and processing. Most application developers have chosen not to take the security issues into account. Security should be a natural part of the application development process. However, implementation of security in handheld device applications is complex and furthermore often tends to distract application developers from developing needed multimedia functionality and to decrease the performance (Havinga, 2000; Ong, Nahrstedt, & Yuan, 2003).

BASIC MOBILE DEVICE SECURITY

Vulnerability Analysis

Security risks associated with mobile devices are almost the inverse of the risks associated with desktop computers. Physical access to desktops can be controlled using physical access control mechanisms in form of guarded buildings and door locks. Electronic access can be controlled using personal firewalls. The physical access control and network intrusion protection mechanisms for desktop computers and wired networks are far away from complete. However, they are considered "good enough" and thus the most serious security concerns for desktop computers are related to data transmission between desktop computers. Mobile devices are small, portable, and thus easily lost or stolen. The most serious security threats with mobile devices are thus unauthorized access to data and credentials stored in the memory of the device.

Physical Access

Mobile phones and PDAs are small portable devices that are easily lost or stolen. Most platforms for mobile devices provide simple software-based login schemes for protecting device access with a password. Such systems can however easily be bypassed by reading information from the device memory without logging in to the operating system. This means that critical and confidential data stored unencrypted in the device memory is an easy target for an attacker who has physical access to the device (Symantec, 2003).

Malicious Software

Smartphones and PDAs have not actually been preferred targets for malware developers until 2004. On the time of writing, malware is still not the most serious security concern. However, the continuous increase of the number of mobile device users worldwide is changing the situation. Current malicious software is mainly focused on Symbian OS and Windows based devices. Malicious software in handheld devices may result in (Olzak, 2005):

- Loss of productivity;
- Exploitation of software vulnerabilities to gain access to recourses and data;
- Destruction of information stored on a SIM card; and/or
- Hi-jacking of air time resulting in increased costs.

Even though malicious software currently does not cause serious threats for the mobile device itself, they cause a threat for a computing network to which the mobile device is connected. Viruses are easily spread to an internal computer network from a mobile device over wireless connections such as infrared, Bluetooth, or WLAN.

Wireless Connection Vulnerabilities

Handheld devices are often connected to the Internet through wireless networks such as cellular mobile networks (GSM, GPRS, UMTS), WLANs, and Bluetooth networks. These networks are based on open air connections and are thus by their nature easy to access. Many Bluetooth networks and especially WLANs are still unprotected. This means that any device, within the network coverage, can access the wireless network. Furthermore, confidential data transmitted over an unprotected wireless network can easily be captured by an intruder.

Security Policy

Examples of security policy rules for mobile device end users as well as for administrators of mobile devices in corporate use are proposed (Taylor, 2004-2005, Part V):

- Make sure that the cell phone, PDA, or smartphone is password protected.
- Use secure remote access VPN to connect to the corporate network for the purpose of checking e-mail.
- Keep a firewall and an anti-virus client with up-to-date anti-virus signatures installed on the handheld if connecting the corporate network.
- Use the security policies on the handheld firewall that are recommended by the corporate security team.

Examples of rules proposed for corporate administrators of mobile devices (Taylor, 2004-2005, Part V):

- End users must first agree to the End-User Rules of Behavior.
- All handheld users are to be setup with a secure remote access VPN client to connect to the corporate network.
- Advice end-users what anti-virus client to use.
- Handheld firewalls are configured to log security events and send alerts to security_manager@company.com.
- Handheld groups and net groups are set up to restrict access privileges only to services and systems required.

Platform Security

The convergence between the fixed Internet and the mobile Internet has raised the question of security of mobile devices. The introduction of packet-based services, such as GPRS and 3G, has opened the wireless industry to new services, but also to new mobile device vulnerabilities. The handsets are actually changing from a formerly closed nature to a more open system. Incidents involving spamming, denial-of-services, virus attacks, content piracy, and malicious attacks have become a growing problem in the wireless world. With this openness a new level of security is required to protect wireless networks and handsets (Sundaresan, 2003).

Security is an integral part of a mobile handset because security affects all parts of the mobile device. Security issues can therefore not be treated as separate add-on features. Instead, security needs to be built into the platform (Trusted Computing Group, 2004).

Mobile Device Forensics

Computer forensics is the process of investigating data processing equipment—typically a PC computer but also mobile device like a PDA computer or a mobile smartphone—to determine if the equipment has been used for illegal, unauthorized, or unusual activities. Computer forensics is especially needed when a mobile device is found or returned after being lost or stolen in order to estimate if and what damage has occurred. By using specially designed forensics software, computer

experts are able to identify suspects and sources of evidence, to preserve and analyze evidence, and to present findings (Phillips, Nelson, Enfinger, & Steuart, 2005; Robbins, 2005).

Physical Protection

Physical security of a mobile device means protection against physical and electromagnetic damage, and protection of stored content against power failures and functional failures. There must be possibilities to recover stored content after damage, after a functional failure, or after an intrusion attack. Furthermore, protection of stored content in case of theft or other loss of the mobile device must be included.

Physical Damage

Most common physical damage is caused by a mechanical shock when the mobile device is dropped or by water when the mobile device has fallen into water or has been exposed to rain.

In case of a mechanical shock the content interface, the display and/or the keyboard, is usually damaged and the stored content is accessible after the content interface is repaired or the electronics and memory modules are transferred to another device. Water will cause damage to the power supply battery and also to electronic components.

Electromagnetic Damage and Unwanted Wireless Communication

The electronic and/or the stored content of a mobile device can be damaged by strong electromagnetic radiation. A switched-on mobile device may also have an always open wireless network connection. For prevention of unwanted electromagnetic radiation and unwanted wireless communication, an electromagnetic shielding bag (MobileCloak™ Web Portal, 2005) can be used.

Power and Functional Failures

The operating system should store configuration changes and sufficiently often, for example every five minutes, backup the working spaces of open applications in non-volatile memory in order to minimize the data losses in case of sudden power failure or other functional failure.

Backup and Recovery

Recovery after power failure or other functional failures is usually managed by the operating system from stored configuration data and from timed backups of the working spaces of open applications. A usual backup function of mobile devices is synchronization with a desktop computer. Examples of synchronization are HotSync in Palm OS, ActiveSync in Windows CE and Windows Mobile, and SyncML in Symbian OS. Recovery from a synchronization backup or from a traditional backup is needed after physical and electromagnetic damage and after an intrusion attack.

Loss and Theft

To minimize damage caused by a lost or stolen device, the following precaution measures are suggested:

- Confidentiality level classification of stored content;
- Content encryption of sensitive data;
- Use of bit wiping software, for stored content with the highest confidentiality level classification; and
- Visible ownership information, for example a phone number to call if a mobile device is found, for return of a lost or stolen mobile device.

Activated bit wiping software will permanently delete data and program code according to user

settings. Good bit wiping software will also delete all related information in working memory (RAM) and in plugged external memory cards. Bit wiping software is typically activated when a wrong device access password is entered a preset number of times and/or the mobile device is not synchronized within a preset timeframe. Because bit wiping software can also be triggered by accident, a fresh synchronization backup or a fresh traditional backup of the stored content should always be available.

Device Access Control

Physical access control mechanisms are ineffective for PDAs and smartphones since such devices are small, portable, and thus easily lost and stolen. Access control mechanisms on the device itself are thus important for protecting the data stored in the device. Currently, there are no widely adopted standards for access control services in mobile devices. Manufacturers are concentrating more on securing the communication protocols than securing stored data in the device. However, this trend is expected to change in the near future. The use of handheld devices is constantly growing, and thus the need for secure access control mechanisms is also increasing.

Access control on a mobile device can be implemented with a combination of security services and features (Perelson & Botha, 2004):

- Authentication service
- Confidentiality service
- Non-repudiation service
- Authorization

Primarily, the security services for access control are authentication and authorization services. The principle of access control on a mobile device is shown in Figure 1.

Authentication

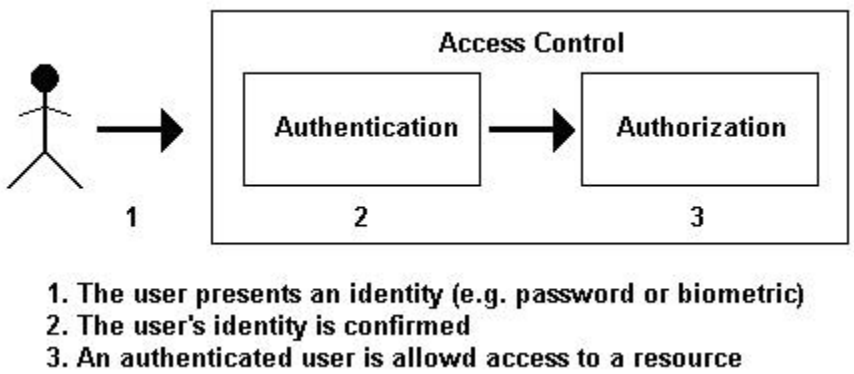
An authentication service is a system for confirming a claimed user identity. There are many methods in which a user is able to authenticate to a handheld device. These methods include:

- Passwords/PINs
- Visual and graphical login
- Biometrics

PIN and Password Authentication

Many handheld devices use a PIN code for user authentication. The PIN is four digits of length and is entered by the user from a ten-digit (0-9)

Figure 1. The principle of access control



numerical keypad. PINs are suitable for personal handheld devices, not containing critical information. However, PINs do not provide sufficient security for handheld devices in corporate use. PINs may be susceptible to shoulder surfing or to systematic trial-and-error attacks due to their limited length and alphabet. Passwords are more secure than PINs since they support a larger alphabet and increase the number of digits in the password string (Jansen, 2003).

Most PDA and smartphone operating systems provide inherent support for traditional alpha-numeric passwords. Strong password authentication solutions can also be implemented by installing additional security software, see section, *Available Security Solutions.*

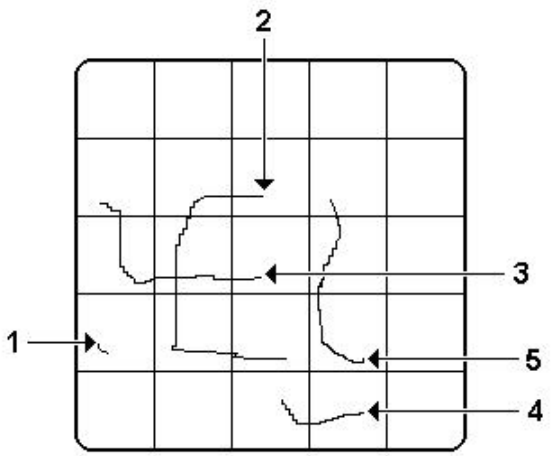
Visual and Graphical Login

Visual authentication allows a user to select a sequence of icons or photo images as a password value. Instead of alpha-numeric characters a user must remember image sequences from a list of pictures shown on the display of the device. A visual authentication method is more user-friendly and more secure than standard password authentication. In order for a password to be secure, it must consist of many digits including both upper and lower case characters as well as both characters and numbers. Picture combinations are for a user easier to remember than complex password strings. Furthermore, a picture password system can be designed to require a sequence of pictures or objects matching a certain criteria and not exactly the same pictures (Duncan, Akhtari, & Bradford, 2004).

Graphical login relies on the creation of graphical images to produce a password value. A user enrolls a password value by drawing a picture on a display grid including block text or graphical symbols. Login process strokes can start anywhere and go in any direction on the touch screen, but they must occur in the same sequence as in the enrolled picture. The system maps each continuous stroke to a sequence of coordinate pairs by listing all cells through which the stroke passes and the order in which the stroke crosses the cell boundaries. An example with a five-stroke password entry, which can be drawn in eight different ways, is shown in Figure 2. The authentication system controls the ordering of the strokes and the beginning and end point of each stroke. The numbered items in Figure 2 indicate in which order the strokes were drawn and point to the starting point of each stroke in the enrolled picture (Jansen, 2003).

Currently, there are no PDAs or smartphones providing inherent support for visual or graphi-

Figure 2. A five-stroke graphical password



cal login. To implement such an authentication method, an additional security application must be installed. Examples of such applications are presented in section, *Available Security Solutions.*

Biometrics

Biometric user authentication is a hardware solution for examining one or more physical attributes of an authorized user. Biometric controls are becoming more and more common in handheld devices (Perelson & Botha, 2004).

Fingerprint Verification

Fingerprint verification provides both user friendliness and security. A user only needs to touch a certain point of the handheld device in order to authenticate to it. The user does not need to remember a complex password string or an image sequence and the security level is still high since every fingerprint is unique.

Signature Verification

Signature verification is a technology where a number of dynamic characteristics from a physical signing process are captured and compared. Dynamic characteristics are speed, acceleration, direction, pressure, stroke length, sequential stroke pattern, and the time and distance when the pen is lifted away from the surface. The physical signature is made with an electronic pen on the touch screen (Jansen, 2003).

Voice Verification

In voice verification identifying and authenticating is based on the user's voice. The enrollment process is normally performed in a way that the user speaks a set of specific words. Usually this process is repeated several times. A template is then extracted from this voice input. This template defines the characteristics of the recorded voice. During authentication, the system prompts the user to pronounce a set of randomly chosen digits as they appear on the display of the handheld device (Jansen, 2003).

Voice verification for smartphones and PDAs has not yet broken through, but a number of companies have recently developed new biometric software and devices (Kharif, 2005).

Authorization

Handheld devices are typically personal and the authentication process infers that the user is authorized. It is often assumed that all data stored on a device is owned by the user. Today, there are large gaps in the features of the authorization services in both smartphone and PDA devices (Perelson & Botha, 2004).

It is becoming more common that handheld devices replace desktop and notebook computers in companies. A single device may be used by several employees and may contain confidential company information. Thus, the need for proper user authorization services is becoming more important. Such services are Perelson and Botha (2004):

- **File Masking:** Certain protected records are prevented from being viewed by unauthorized users.
- **Access Control Lists:** A list defines permissions for a particular object associated with a user.
- **Role-Based Access Control:** Permissions are defined in association with user roles.

Storage Protection

Storage protection of a mobile device includes online integrity control of all stored program code and data, optional confidentiality of stored user data, and protection against unauthorized tampering of stored content. Protection should

include all removable storage modules used by the mobile device.

The integrity of the operating system code, the program code of installed applications, and system and user data can be verified by using traditional tools like checksums, cyclic redundancy codes (CRC), hashes, message authentication codes (MAC, HMAC), and cryptographic signatures. Only hardware-based security solutions for protection of verification keys needed by MACs, HMACs, and signatures provide strong protection against tampering attacks, since a checksum, a CRC, and a hash of a tampered file can easily be updated by an attacker. Online integrity control of program and data files must be combined with online integrity control of the configuration of a mobile device. This is needed to give sufficient protection against attempts to enter malicious software like viruses, worms, and Trojans.

Required user data confidentiality can be granted by file encryption software. Such software also protects the integrity of the stored encrypted files, since successful decryption of an encrypted file is also an integrity proof.

Network and Network Service Access Control

Once a user is authenticated and granted access to the handheld device, the device can be connected to and used in several types of networks and network services. This section presents access control mechanisms and user authentication issues related to some of these systems.

Hardware Authentication Tokens

Hardware tokens are used in handheld devices to authenticate mobile end users to mobile networks and network services. Such hardware tokens are SIM (subscriber identity module), PKI SIM (public key infrastructure SIM), USIM (universal SIM) and ISIM (IP multimedia services identity module) chips.

SIM

A basic SIM card is a smartcard securely storing an authentication key identifying a GSM network user. The SIM card is technically a microcomputer, consisting of a CPU, ROM, RAM, EEPROM, and I/O (input/output) circuits. This microcomputer is able to perform operations, such as cryptographic calculations with the individual authentication key needed for authenticating the subscriber. The SIM card also contains storage space for SMS (short message services) messages, MMS (multimedia messaging system) messages, and a phone book. The use and content of a SIM card is protected by using PIN codes.

PKI SIM

A PKI SIM card is a basic SIM card with added PKI functionality. A co-processor is added to take care of decryption and signing with private keys. A PKI SIM card contains space for storing private keys for digital signatures and for decryption (Setec Portal, 2005).

PKI SIM cards open the possibility for handheld users to generate digital signatures and authenticate to online services. Currently, the use of PKI SIM cards is not widely supported. However, systems for utilizing this technique are being developed.

USIM

A USIM card is a SIM used in 3G mobile telephony networks, such as UMTS. USIM is based on a different type of hardware; it is actually an application running on a UICC (universal integrated circuit card). The USIM stores a pre-shared secret key as the basic SIM.

ISIM

An ISIM card consists of an application (ISIM) residing on a UICC. The ISIM application provides

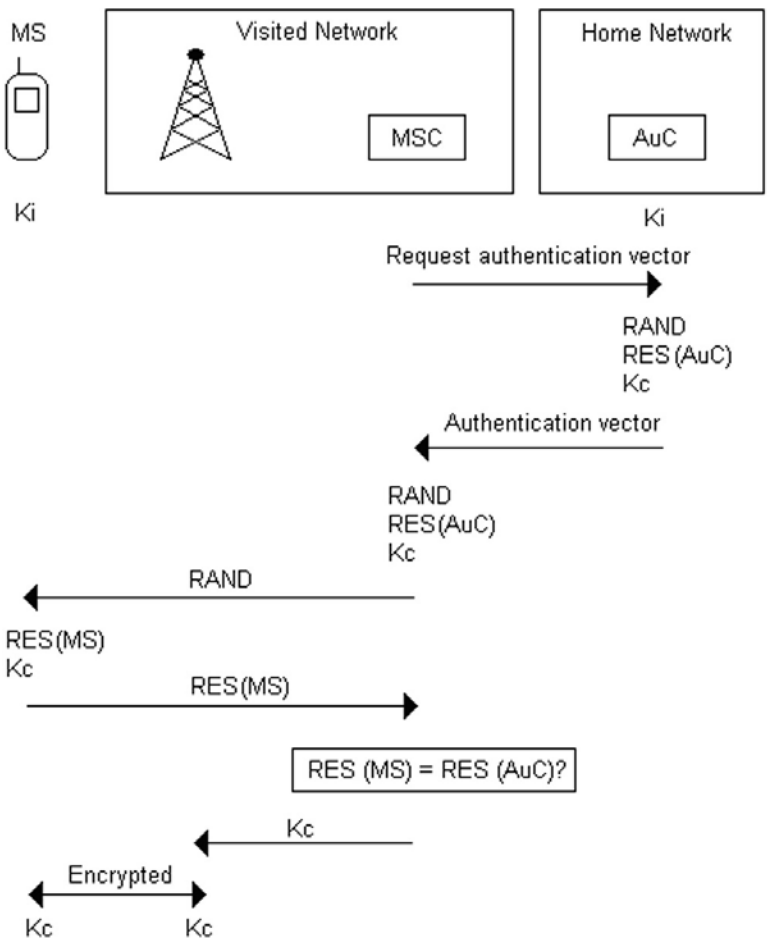
secure authentication of handheld users to IMS (IP multimedia systems) services (Dietze, 2005).

Cellular Networks

GSM/GPRS

User authentication in GSM networks is handled by a challenge-response based protocol. Every MS (mobile station) shares a secret key Ki with its home network. This key is stored in the SIM card of the MS and the AuC (Authentication Centre) of the home network. Ki is used to authenticate the MS to the visited GSM network and for generating session keys needed for encrypting the mobile communication. The authentication process, shown in Figure 3, is started by the MSC (Mobile Switching Centre) which requests an authentication vector from the AuC of the home network of the MS. The authentication vector, generated by the AuC, consists of a challenge response pair (RAND, RES) and an encryption key Kc. The MSC of the visited network sends the 128-bit RAND to the MS. Upon receiving the RAND, the MS computes a 32-bit response (RES) and an encryption key Kc using the received RAND and the Ki stored in the SIM. The MS sends the RES back to the MSC. MSC verifies the identity of the MS by comparing the received RES from the MS with the received RES from the AuC. If they

Figure 3. GSM authentication and key agreement



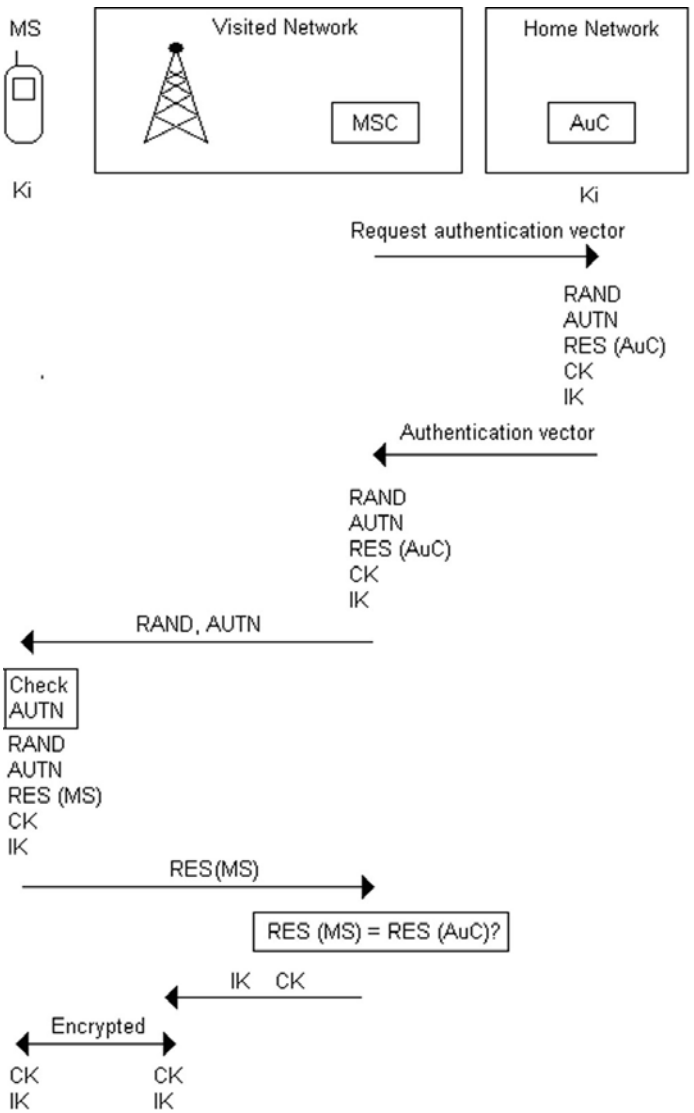
match, authentication is successful and the MSC sends the encryption key Kc to the base station serving the MS. Then the MS is granted access to the GSM network service and the communication between the MS and the base station is encrypted using Kc (Meyer & Wetzel, 2004).

GSM networks provide reasonably secure access control mechanisms. The most serious concern is the lack of mutual authentication. This opens a possibility for an attacker to setup a false base station and imitate a legitimate GSM network. As a result, the key Ki can be cracked and the attacker can then impersonate a legitimate user (GSM Security Portal, 2005).

UMTS

The authentication and key management technique used in UMTS networks is based on the same principles as in GSM networks (see Figure 4). A secret authentication key is shared between the network and the MS. This key is stored on the MS USIM

Figure 4. UMTS authentication and key agreement



and in the AuC of the home network. Unlike in GSM networks, UMTS networks provide mutual authentication. Not only MS is authenticated to the UMTS network but also the UMTS network is authenticated to MS. This protects MS from attackers trying to impersonate a valid network to the MS. Network authentication is provided by a so-called authentication token AUTN.

The MSC (Mobile Switching Centre) of the visited network sends the AUTN together with the authentication challenge to MS. Upon receiving the AUTN, containing a sequence number, the MS checks whether it is in the right range. If the sequence number is in the right range the MS has successfully authenticated the network and the authentication process can proceed. The MS computes an authentication response, here called RES, and encryption and integrity protection keys, called CK and IK, and sends these back to the MSC. The MSC verifies the identity of the MS by checking the correctness of the received RES. Upon successful authentication, the MSC sends the encryption key CK and integrity key IK to the UMTS base station. The MS is now able to communicate with the UMTS network and the communication between the MS and the base station is encrypted with CK and the integrity is protected with IK (Meyer & Wetzel, 2004).

Wireless Personal Area Networks

IrDA

The IrDA standard does not specify any security measures. However, since IrDA is a short line of sight-based connection, security threats can be eliminated by physical security measures.

Bluetooth

Bluetooth technology provides device authentication, not actually user authentication. The whole authentication process can be divided into two phases, an initial process called pairing and a mutual device authentication process.

Pairing

Two Bluetooth devices, that want to set up a connection, share a PIN code which is entered on both devices in the beginning of the pairing process. The PIN code, the address of the Bluetooth device, and a 128-bit random number is used as inputs to an algorithm (E1) for creating an initialization key. A new random value is generated on both devices and exchanged after XORing it with the initialization key. With this new 128-bit random value, and the Bluetooth device address, the common shared secret key, called link key, is generated using the E21 algorithm (Gehrmann et al., 2004; Xydis & Blake-Wilson, 2002).

Mutual Authentication

After pairing, the actual device authentication is performed through the use of a challenge-response scheme. One of the devices is the verifier and the other is called a claimant. The verifier generates a 128-bit challenge value. The claimant then uses the algorithm E1 with the challenge, its 48-bit Bluetooth address, and the link key as inputs, for creating a 128-bit value. The 32 most significant bits of this value are returned to the verifier. The verifier verifies the response word by performing the same calculations. If the response value is successful, then the verifier and the claimant change roles and repeat the authentication process (Gehrmann et al., 2004; Xydis et al., 2002).

Authorization

Bluetooth technology also offers a way of performing user authorization. Devices have two levels of trust. They are divided into trusted or untrusted devices. A device is trusted only when it first has been authenticated. Interplay of authentication and authorization defines three service levels (see Table 1) (Gehrmann et al., 2004).

At service level 3, any device is granted access to any service. At the next level, only authenticated

Table 1. Bluetooth service levels

	Authorization	Authentication	Encryption
Service Level 1	Yes	Yes	Yes
Service Level 2	No	Yes	Yes
Service Level 3	No	No	Yes

devices get access to all services, and at service level 1, only authenticated devices are granted access to one or more certain services.

Security Analysis

One of the major weaknesses in Bluetooth access control is the lack of support for user authentication. A malicious user can easily access network resources and services with a stolen device. Furthermore, PIN codes are often allowed to be short and thus susceptible to attacks. However, the coverage range of a Bluetooth network is very short. Malicious access to a Bluetooth network can therefore mostly be prevented by the use of physical access control measures.

Wireless Local Area Networks

Implementation and use of secure access control mechanisms is essential in order to protect WLANs from unauthorized network access. WLANs became, during their first years, known for their serious security vulnerabilities. One of the most significant concerns has been the lack of proper user authentication methods. Today, WLANs provide acceptable security through the recently ratified security standard IEEE 802.11i.

Access Control Based on IEEE 802.11

The authentication mechanisms defined in the original WLAN standard IEEE 802.11 are weak and not recommended. The standard only provides device authentication in form of the use of static shared secret keys, called WEP (wired equivalent privacy) keys. The same WEP key is shared between the WLAN access point and all authorized clients. WEP keys have turned out to be easily cracked. If a WEP key is cracked by an intruder, the intruder gets full access to the WLAN.

WEP authentication can be strengthened by using MAC filters and by disabling SSID broadcasting on the access point. However, SSIDs are easily determined by sniffing probe response frames from an access point and MAC address are easily captured and spoofed.

Access Control Based on IEEE 802.11i

The recently ratified WLAN security standards WPA and WPA2 address the vulnerabilities of WEP. WPA is a subset of the IEEE 802.11i standard, and WPA2 provides full 802.11i support. The difference between WPA and WPA2 is the way how the communication is encrypted. Furthermore, WPA2 provides support for ad-hoc networks which is missing in WPA. User authentication in WPA and WPA2 are based on the same techniques. Currently, there are already a number of PDAs and smartphones supporting WPA. WPA2, however, is still not supported in any handheld device (Wi-Fi Alliance Portal, 2005).

Pre-Shared Key. 802.11i provides two security modes: home mode and enterprise mode. The home mode is based on a shared secret string, called PSK (pre-shared key). Compared to WEP, PSK is never used directly as an input for data encryption algorithms. For large WLANs, the enterprise mode is recommended. The home mode is suitable for small WLAN environments, such as small office and home WLANs, where the number of users is small.

IEEE 802.1X. The enterprise security mode utilizes the IEEE 802.1X standard for user authentication. The 802.1X standard is designed to address open network access. Three different components are involved: supplicant (client), authenticator (WLAN access point) and AAA (authentication, authorization, and accounting) server. The supplicant wants to be authenticated and accesses the network via the authenticator. The AAA server, typically a RADIUS (remote authentication dial-in user service) server, works as a back-end server providing authentication service to an authenticator. The authentication server validates the identity and determines, from the credentials provided by the supplicant, whether the supplicant is authorized to access the WLAN or not.

The authenticator works as an intermediary between the supplicant and the authentication server passing authentication information messages between these entities. Until the supplicant is successfully authenticated on the authentication server, only authentication messages are permitted between the supplicant and the authentication server through the authenticator's uncontrolled port. The controlled port, through which the supplicant can access the network services, remains in *unauthorized state* (see Figure 5). As a result of successful authentication, the controlled port switches to *authorized state*, and the supplicant is permitted access to the network services (see Figure 6).

EAP (Extensible Authentication Protocol). 802.1X binds the EAP protocol which handles the exchange of authentication messages between the supplicant and the AAA server. The exchange is performed over the link layer, using device MAC addresses as destination addresses. A typical EAP authentication conversation between a supplicant and an AAA server is shown in Figure 7.

EAP supports the use of a number of authentication protocols, usually called EAP types. The following EAP types are WPA and WPA2 certified (Wi-Fi Alliance Portal, 2005):

- EAP-TLS (EAP-Transport Layer Security);
- EAP-TTLS (EAP-Tunneled Transport Layer Security);
- PEAPv0/EAP-MSCHAPv2 (Protected EAP version 0/EAP-Microsoft Challenge Authentication Protocol version 2);
- PEAPv1/EAP-GTC (PEAPv1/EAP-Generic Token Card); and
- EAP-SIM.

Figure 5. 802.1X authentication in unauthorized state

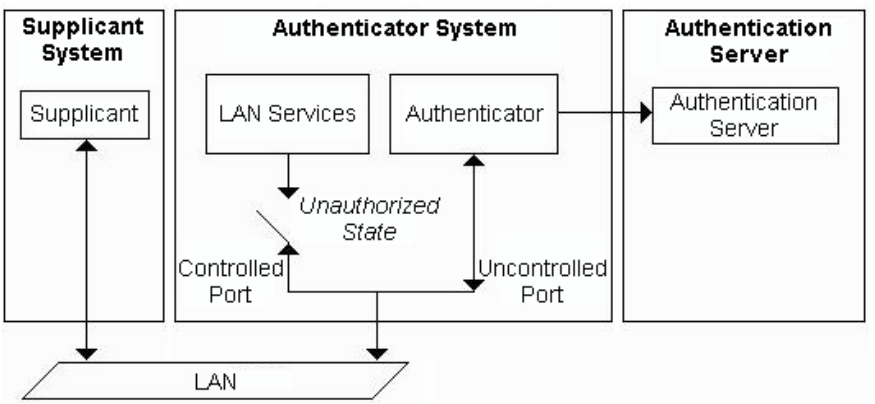


Figure 6. 802.1X authentication in authorized state

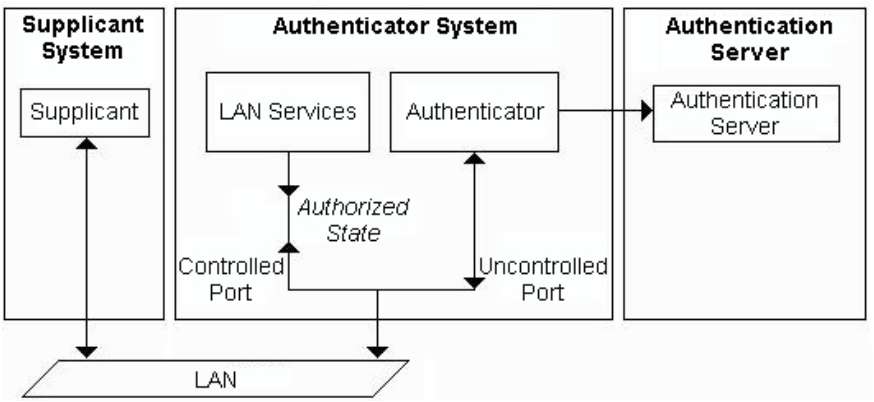
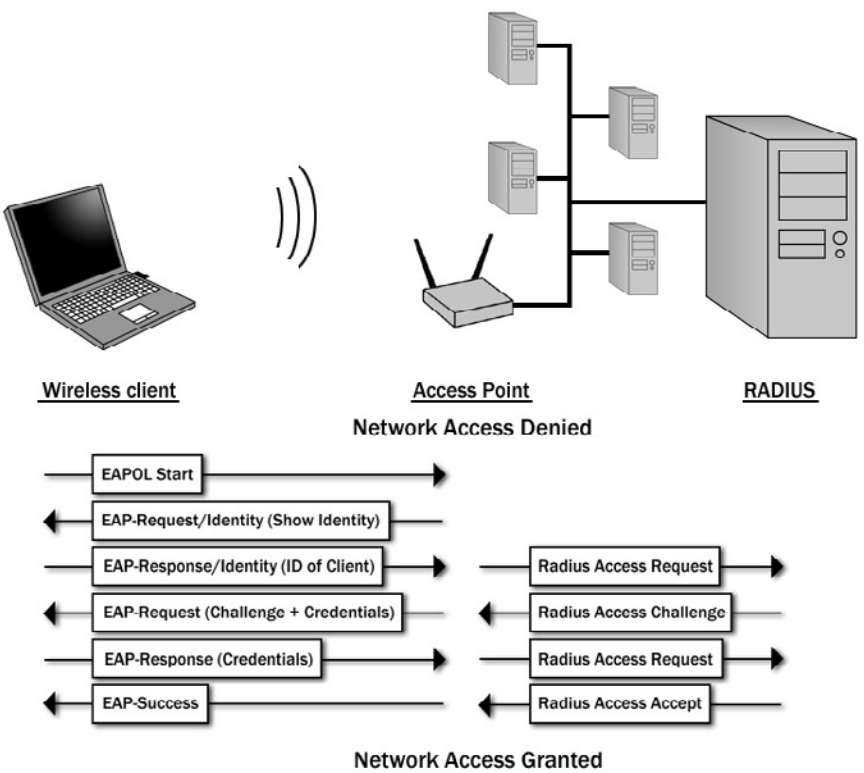


Figure 7. EAP authentication exchange messages



EAP-TLS, EAP-TTLS, and EAP-PEAP are based on PKI authentication. EAP-TTLS and EAP-PEAP however only use certificate authentication for authenticating the network to the user. User authentication is performed using less complex methods, such as user name and password. EAP-TLS provides mutual certificate based authentication between clients and authentication servers. Therefore, a X.509 based certificate is required both on the client and on the authentication server for user and server authentication (Domenech, 2003).

EAP-SIM is an emerging EAP authentication protocol. This standard is still an IETF draft. EAP-SIM is based on the existing GSM mobile phone authentication system. A user is able to authenticate to the network using the secret key and algorithms embedded on the SIM card. In order to implement EAP-SIM, a RADIUS server supporting EAP-SIM and equipped with a GSM/MAP/SS7 (GSM/Mobile Application Part/Signalling System 7) gateway is needed. Additionally, the client software must support the EAP-SIM authentication protocol. The RADIUS server contacts the user's home GSM operator through the GSM/MAP/SS7 gateway and retrieves the GSM triplets. The triplets are sent to the client via the access point, and if the supplicant and the user's SIM card are able to validate the GSM triplets, then the RADIUS server requests the access point to grant the client network access.

Network Connection Security

Connection security means:

- Integrity of communication;
- SIM, USIM, PKI SIM, ISIM cards in cellular mobile networks (GSM, GPRS, UMTS);
- Confidentiality of communication (SSH, VPN, SSL, voice, video);
- Intrusion prevention
    - traffic filtering (firewall)
    - firewall log of connection attempts;
- Malware rejection: Anti-virus software and anti-spyware;
- Mutual authentication of communicating partners;
- Security settings and commitment to security policy rules controlled by centralized security management software;
- Control of remote synchronization (Palm OS/HotSync, Windows CE & Windows Mobile/ActiveSync, Symbian OS/SyncML based);
- Security audits based on communication event logging, analysis of logged information, alerts, and alarms; and
- Shielding the mobile device from unwanted wireless communication.

Basic Communication Security

Basic communication security of a mobile device can be defined as intrusion prevention and malware rejection. The basic intrusion prevention tool is a configurable firewall with communication event logging and alert messaging. The basic malware rejection tools are anti-virus and anti-spyware with suspicious event alarming features. The core of a malware rejection tool is a malware recognition database. Malware rejection tool providers constantly update this database and the updated malware recognition database is available to users through some network connection. An installed malware rejection tool should always use the latest update of the malware recognition database. Anti-virus software for mobile devices is delivered for example by Symantec Corporation (2005) and Kaspersky (2007).

Authentic Data Communication

Authentic data communication is based on mutual authentication of communicating parties. In 2G cellular networks (GSM Data) authentication is uni-directional. The mobile device is authenticated to the cellular network by use of the shared secret

key in the SIM card. Mutual authentication, for example, based on a public key certificate, is however possible for packet data communication in GSM networks (GPRS) in addition to PIN-based GSM authentication. In 3G cellular networks, like UMTS, authentication is mutual. The mobile device and the network are authenticated to each other by the authentication and key agreement (AKA) mechanism (Cremonini, Damiani, De Capitani di Vimercati, Samarati, Corallo, & Elia, 2005).

In a WLAN authentication is mutual for WPA and IEEE 802.11i (WPA2). The authentication of a mobile client is based on presented credentials and information registered in an AAA server. The authentication protocol, EAP, also requires authentication of the AAA server to the mobile client (Pulkkis, Grahn, Karlsson, Martikainen, & Daniel, 2005). Also a Bluetooth connection can be configured for mutual authentication. The default security level of a Bluetooth service is:

- **Incoming Connection:** Authorisation and authentication required.
- **Outgoing Connection:** Authentication required (Muller, 1999).

Integrity and Confidentiality of Data Communication

Confidentiality and integrity of all data communication to and from cellular mobile networks (GSM, GPRS, UMTS) is provided by the security hardware of SIM/USIM/PKI SIM/ISIM cards in mobile devices.

For data communication through other network types (WLAN, Bluetooth, IrDA) connection specific security solutions must be installed, configured, and activated. Alternatively, end-to-end security software like VPN and SSH must be used. Available PDA VPN products are listed in (Taylor, 2004-2005, Part IV).

For WLAN connections, available solutions for confidentiality and integrity of all data communication are WEP, WPA, and IEEE 802.11i (WPA2). WEP security is however weak, since WEP protection can be cracked from recorded WEP protected data communication (WEPCrack, 2007).

For Bluetooth connections link-level security corresponding to security mode 3 should be used (Sun, Howie, Koivisto, & Sauvola, 2001a).

Roaming Security

Widely accepted Internet protocol standards supporting roaming for mobile devices are Mobile IP (Perkins, 2002) and SIP (Rosenberg et al., 2002). Mobile IP supports network level roaming, which is transparent for all IP-based network applications. SIP supports application level roaming, mainly for multimedia streaming services. Roaming security means:

- mutual authentication of a mobile device and the access network;
- integrity protection and confidentiality of all data communication, including roaming control signaling in handover situations.

A mobile device with Mobile IP support has an IP address registered by a home agent in the home network. A roaming mobile device visiting a new foreign network is registered with an additional IP address, the care-of-address by a foreign agent in the visited network and also by the home agent. The home agent can then redirect IP packets with the home address of the mobile node to the foreign network, which is presently visited by the mobile node. Home and foreign agent functionality is usually implemented on the gateway router of a network.

Authentication services for the Mobile IP protocol are:

- Security associations for authentication of mobile device registration messages (Rosenberg et al., 2002);

- AAA server based authentication of the home agent and the foreign agent of a mobile device (Calhoun, Johansson, Perkins, Hiller, & McCann, 2005; Glass, Hiller, Jacobs, & Perkins, 2000); and
- Public key cryptography based authentication of a mobile device (Hwu, Chen, & Lin, 2006; Lee, Choi, Kim, Sohn, & Park, 2003).

Secure Mobile IP communication in a handover situation can be achieved:

- on data link level by the WLAN security protocol WPA or IEEE 802.11i, when the new access network is a WLAN;
- as end-to-end protection by the security protocols IPSec, SSL, SSH, or by mobile node identity based public key cryptography (Barun & Danzeisen, 2001; Hwu, Chen, & Lin, 2006; Islam, 2005).

Security of SIP supported mobile device roaming is described in the section, *Mobile Multimedia Security.*

Connection Security Management

Security settings and commitment to security policy rules should be controlled by centralized security management software. Security audits based on communication event logging, analysis of logged information, and on alerts and alarms should be performed with timed and manual options.

Special attention should be paid to control of remote synchronization (Palm OS/HotSync, Windows CE & Windows Mobile/ActiveSync, Symbian OS/SyncML based). Remote synchronization should be disabled when not needed. Desktop and laptop computers should have basic communication security features like personal firewalls, anti-malware protection software with updated malware recognition data and latest software patches installed.

Use and attempts to use remote synchronization ports (see Table 2) should be logged and alerts and alarms should be triggered by unauthorized use or usage attempts. Passwords used by synchronization software in desktop and laptop computers should resist dictionary attacks and the PC/Windows option to save connection passwords should not be used.

SECURITY OF MULTIMEDIA NETWORK APPLICATIONS

Security requirements for multimedia network applications and approaches to fulfill these requirements are surveyed in Sun, Howie, Koivisto, and Sauvola (2001b). The requirements include authentication, availability, protection, and performance.

Authentication Requirements

Authentication requirements for secure multimedia network applications are source authentication, receiver authentication, and authentication of delivered content.

Receiver authentication means that a user of a multimedia service must identify himself/herself with a password, by proving the possession of a

Table 2. TCP and UDP ports used by synchronization software

	TCP Ports Used	UDP Ports Used
ActiveSync	990, 999, 5678, 5679	
HotSync	14237, 14238	14237
SyncML based	80 (http)	

security token, or biometrically before he/she is authorized for the service. Source authentication usually means that a user of a multimedia service checks the possession of a security token for the server or the provider before using the service. Digital signing and watermarking techniques can be used for authentication of delivered content.

A digital watermark is an imperceptible signal, which can be inserted into a digital content—a document, an image, an audio file/stream, a video file/stream, and so forth—for purposes like copyright control and binding digital content to content owners. Watermarks resist tampering and are robust to signal transformations like file filtering and compression. Content copyright and ownership can be checked with special watermark reading software. Digital pirates can be forensically revealed by watermark checks. A special type of watermarks, called annotation watermarks, can be used for access control based on binding usage rights to digital content. However, the current state of watermarking technology is imperfect. A hacker knowing the used watermark technique can often tamper watermarks beyond recognition (Cox & Miller, 1997; Dittmann, Wohlmacher, & Ackermann, 2001; Liu, Reihaneh, & Sheppard, 2003; Stamp; 2002).

Availability Requirements

Delivery of multimedia content to authorized receivers must be granted. Granted delivery of streaming multimedia implies quality of service (QoS) requirements, which include sufficient bandwidth from source to receiver for successful real-time content transport.

Requirements on IP for granted content delivery to authorized receivers are secure routing and DNS security. A misconfigured, faulty, or tampered router misdirects multimedia content transport and a fake DNS gives an incorrect IP address of the receiver.

Protection Requirements

Delivered multimedia content must protected against tampering, capture, unauthorized copying, and denial-of-service (DoS) attacks. With methods based on encryption, digital signing, and watermarking, some protection can be achieved.

Performance Requirements

Multimedia network applications usually generate and transmit a waste amount of data. Some multimedia applications, like interactive multipoint video streams, must also meet real-time constraints. Cryptographic operations needed by security services are therefore associated with strict execution time constraints, which can be met only by hardware implemented cryptographic algorithms. Also the complexity and specific characteristics of video coding technologies require careful choice as well as special design and application of cryptographic algorithms in necessary security services.

Secure Multimedia Streaming Services

For streaming service sessions both the session setup SIP signaling and the media transport should be protected. Security mechanisms in SIP for authentication of SIP user agent clients and confidential end-to-end and hop-to-hop communication are surveyed in Salsano, Veltri, and Papalilo (2002).

Streaming content delivery can be protected with the Secure Real-time Transport Protocol (SRTP) (Baugher, McGrew, Naslund, Carrara, & Norrman, 2004), which can provide confidentiality, message authentication, and reply protection to the RTP traffic and to the control traffic for RTP, the Real-time Transport Control Protocol (RTCP), which provides out-of-band control information

for an RTP flow (Schulzrinne et al., 2003). RTPC cooperates with RTP in the delivery and packaging of multimedia data, but the protocol does not actually transport any data itself. The function of RTCP is to gather statistics on a media connection and also information such as bytes sent, packets sent, lost packets, jitter, feedback, and round trip delay. An application may then use this information to increase the QoS. SRTCP (secure real-time transport control protocol) is a security protocol designed to provide the same security-related features to RTCP, as the ones provided by SRTP to RTP.

RTP does not provide any QoS guarantee. QoS can, however, be achieved using resource reservation techniques in the network. An example of such a technique is defined in RSVP (resource reservation protocol) (Braden, Zhang, Berson, Herzog, & Jamin, 1997; Featherstone & Zhang, 2004).

The RTP protocol itself is not suitable for integrated heterogeneous network environments consisting of both wired and wireless networks. As a solution for this weakness, mixers have been added to the RTP solution. The mixers are using a process called transcoding in order to deal with different levels of bandwidth. The transcoding process translates an incoming data stream into a different one with possible lower data-rate.

The use of mixers and translators to perform transcoding processes to achieve mobility support compromises security for example in video conferencing systems. Mixers and translators must decrypt data in order to be able to manipulate the data. After manipulation, the data streams are again re-encrypted. Thus, end-to-end confidentiality is compromised and furthermore since the data has been altered, it is impossible to perform source authentication.

The RSVP provides an authentication technique called hop-by-hop authentication (Baker, Lindell, & Talwar, 2000). This technique has however a high overhead, and is thus not suited for use in wireless networks with limited bandwidth. The hop-by-hop authentication in RSVP does furthermore not provide any key management or confidentiality services. This means that a separate key distribution algorithm is needed prior to the start of a conference and an attacker, who is eavesdropping the resource reservation messages could infer useful and sensitive information.

A solution to address security concerns with RSVP is proposed in RSVP-SQoS (RSVP with scalable QoS protection) (Talwar & Nahrstedt, 2000). This solution makes the design scalable by dividing a network into different sub-networks and it assumes that security requirements within one sub-network are weaker than the requirements for inter-sub-networks. Immediately, when a security violation is detected, the resource reservation is removed. However, considering mobile multimedia streaming services, this solution is not suitable since user mobility is not taken into account.

An end-to-end authentication scheme for resource reservation designed to eliminate DoS attacks, where an attacker alters the amount of resource requests, is worked out in Wu, Wu, Huang, and Gong (1999). In this solution, a digitally signed request is sent to the receiver by the source of the data stream. The digital signature is generated with the source's private key. If the receiver is able to verify the signed request, then it sends out a resource reservation message, which is a piggybacked version of the original request from the source, and a resource reservation request. These are also digitally signed before sent. As this message is sent back along the network, intermediate routers check if the piggybacked message is equivalent to the original one and when the message arrives back to the sender, the sender checks the signature of the packet. If any of these checks fail, the resource reservation is rejected. This protocol has, however, two major weaknesses. The first one is that it does not provide confidentiality to the reservation message. A potential attacker can thus eavesdrop on the messages. The second major weakness is that it lacks any support for mobility.

MOBILE MULTIMEDIA SECURITY

A mobile device for multimedia applications must fulfill the specific security requirements of mobile multimedia in addition to:

- all basic security requirements of mobile data communication; and
- all security requirements of multimedia network applications.

Security requirements of multimedia processing are included in the framework model of mobile security in Zeng, Zhuang, and Lan (2004).

A mobile device for multimedia applications should support:

- SIP signaling based application level mobility for multimedia streaming services; and
- Mobile IP-based network level mobility for file download services and carousel services for text and images.

Mobil IP functionality with security mechanisms described in the section, *Network Connection Security*, as well as SIP user agent functionality with security mechanisms surveyed in Salsano et al. (2002) should thus be integrated in secure mobile devices for multimedia applications. SIP signaling can during inter-domain roaming be separated from other network traffic by domain edge routers (Nasir & Mah-Rukh, 2006).

SRTP (Baugher et al., 2004) is used to encrypt the video, audio, or data stream for securing a multimedia streaming service, since real-time media transport is based on RTP and UDP. Setting up a secured session with SIP signaling involves exchange of separate SRTP key between the mobile device and the host delivering the stream for the specific media type (Dutta et al., 2004a).

Mobile Multimedia Device Security Based on IPSec VPN Technology

Mobile device security based on a IPSec VPN architecture is proposed for mobile multimedia services in a MAN network in Zhang, Lambertsen, and Yamada (2003). Bandwidth requirements of multimedia services are provided by a PON (passive optical network) backbone access network, which allows IP multicasting through optical fibres to a group of adjacent wireless base stations implemented by optical network units (ONU). The PON gateway to the Internet is an optical line terminal (OLT). An IPSec VPN agent with a VPN tunnel to the home network of the mobile devices is installed on this OLT. A mobile device entering the MAN is registered by the VPN agent, which also records the multicast address a roaming domain consisting of some adjacent ONU base stations. The location information of the mobile device is also updated in the VPN server in the home network of the mobile device. Smooth handover of IP multicast communication is now possible even if the mobile device roams rapidly between the ONU base stations. The privacy of the mobile device and the communication between the mobile device and the VPN agent can be protected by another IPSec VPN tunnel. The SAs and the VPN tunnel must not be renegotiated when the mobile device roams to another ONU base station, but the radio channel to the new ONU must be recreated and the location information of the mobile device must be updated at the VPN agent and at two ONUs. A correspondent C can now use the IP address in the home network of a mobile device. The VPN server in the home network can redirect the IP packets to the visited MAN network through the VPN tunnel to the VPN agent, which forwards the packets—unencrypted, pre-shared key encrypted, or ESP encrypted in another VPN tunnel—to the mobile device through the registered ONU.

Secure Mobile Video Conferencing

Video conferencing is one of the proposed future multimedia applications for mobile handheld devices. In a video conferencing system one or more data streams must be delivered to a client device in real time. The data streams may consist of video, audio, or other types of multimedia data. Two or more participants are taking part in a conference and the participants may want to have security provisions to protect their communication. For a mobile video conferencing system, where one or more of the participants are mobile, there are many challenging security issues:

- Implementing security with minimal loss of processing power;
- Implementing integrity and authentication algorithms, where authentication will not fail because of single bit errors caused by disturbances or disconnections in a wireless network;
- Roaming support;
- Protecting video streams from interception and interference; and
- Providing security for multicast multimedia data.

Security solutions for mobile multimedia video conferencing systems are under ongoing research. An example of such a research project is presented in Featherstone and Zhang (2003). The project started with studies of limitations and requirements for implementing security in mobile video conferencing environments. Related work was investigated to find out how current solutions can be improved to meet the requirements.

The requirements for mobile video conferencing can, according to the research project presented in Featherstone and Zhang (2003), be divided into three categories:

- Performance requirements
- Functional requirements
- Security requirements

Performance Requirements

Performance requirements of a mobile video conferencing system include QoS (quality of service) provision, minimal overheads, and accommodating heterogeneous network technologies. Real-time data needs bounded response times. Thus, bandwidth always needs to be guaranteed, also when a user is roaming between two mobile networks. A video conferencing system with implemented security and QoS provision introduces extra overhead to the system and also consumes extra bandwidth. These parameters need to be minimized in order to maintain QoS provision. Current and also future communication systems integrate both wired and wireless access networks of different standards. Thus, a video conferencing system design must be independent of underlying network technologies.

Functional Requirements

A video conferencing system should include at least the following functions:

- Initiation of a conference;
- Finishing of a conference;
- Admitting people to a conference;
- Allow someone to leave a conference;
- Deal with unexpected departure of a member from a conference;
- Split a conference;
- Merge two or more conferences; and
- Temporarily suspend a member from a conference.

Security Requirements

Security requirements in a mobile video conferencing system include:

- Confidentiality of the multiple multimedia data streams generated by a conference system;

- Integrity of multimedia data;
- Authentication of each member of a conference;
- Anonymity service for preventing outsiders from knowing that a conference is taking or has taken place;
- Non-repudiation service for providing evidence that a particular conference has actually taken place;
- Efficient and effective key exchange process for providing key distribution in a multicast environment;
- End-to-end protection; and
- Flexible security services providing users the possibility to choose different levels of protection.

CURRENT MOBILE MULTIMEDIA RESEARCH

The purpose of this section is to provide a picture of current and future trends in mobile multimedia security research. The research area is wide, and there is a large amount of ongoing research projects. Therefore, this section will only concentrate on a few important areas, such as secure roaming and video conferencing. A few examples of related research projects are presented. A testbed for developing and testing secure mobile multimedia systems is also presented.

Proposals for Secure Roaming

The current trend in mobile networking is towards mobile devices having multiple network interfaces such as WLAN, UMTS, and GPRS. A mobile user may want to securely and continuously access an enterprise network as well as securely receive real-time data such as multimedia data streams regardless of whether the user is physically located in the enterprise or not. The user should also be able to seamlessly roam between different types of wireless networks and still maintain an one-going secure application session. In order for a user to access an enterprise network from an external network it is typically required to use VPN technique. Using VPN, the enterprise network is able to authenticate the user and to determine whether the user should be permitted to use enterprise network applications or not. A key issue in developing a mobile multimedia system supporting seamless and secure roaming across heterogeneous radio systems is how to fulfill the security requirements and at the same time maintain QoS.

Current VPN technologies, such as IPSec, do not have sufficient capabilities to support seamless mobility. For instance, an IPSec tunnel will break when a mobile device changes its IP address as a result of roaming between two networks. Researchers are currently developing a new version of IKE (Internet Key Exchange), IKEv2, also known as MOBIKE. This version adds a mobility extension to the IKE protocol. The current MOBIKE specification is available as an IETF Internet-draft (Kivinen & Tschofenig, 2006). However, the work is still in a quite early stage. Another problem with VPN and mobility is that the VPN establishment requires manual actions from users when a time-variant password is used to set up the VPN. Furthermore, a VPN may cause significant overhead.

Roaming Based on Mobile IP and SIP

A secure universal mobility (SUM) network has been proposed (Dutta et al., 2004a). The network is able to support seamless mobility and the mobile user does not need to maintain an always-on VPN session. The SIP protocol can be used to achieve dynamic VPN establishment and Mobile IP for mobility support. A dynamic VPN is achieved by using a double tunneled Mobile IP. This Mobile IP system needs two home agents, an internal home agent, and an external home agent. One tunnel is set up from the mobile device to the external home agent and another tunnel is set up from the

external home agent to the internal home agent. When both tunnels are up, a corresponding node or a mobile user is able to initiate a communication by sending a SIP INVITE message. The receiver then checks if there is an existing VPN session. If not, IKE is used to negotiate security parameters and establish a new VPN session.

A testbed realization of secure universal mobility is also presented. Testbed experiments proved that smooth handoff is achieved during roaming between heterogeneous networks. However, an additional delay appeared while moving from a 802.11 network to a cellular network. VoIP and video streaming traffic was also tested.

Host Identity Protocol

When the current Internet architecture was designed, functionalities such as mobility and multi-homing were not taken into account. An IP address represents both an identity and a topological location of the host. This overloading has led to several security problems, such as the so-called address ownership problem (Nikander, 2001). This makes IP mobility and multi-homing hard from the security point of view.

A new protocol, HIP (host identity protocol) is currently under development and the current specifications are available as an IETF draft (Moskowitz & Nikander, 2006). In the HIP architecture, the locators and end-point identifiers are separated from each other at the network layer of the TCP/IP stack. A new cryptographic name space, the HI (host identity) is introduced, which is typically a cryptographic public key. This key serves as the end-point identifier of a network node. The IP address retains the role of a locator. Each host has at least one HI assigned to its networking kernel or stack. The HIs can be either public or anonymous. Public HIs may be stored in directories, such as DNS, allowing the host to be contacted by other hosts. In other words, HIP provides a binding between the HIs and the IP addresses using DNS.

The major advantages of the HIP architecture are that the problems of dynamic readdressing, anonymity, and authentication are solved. In HIP, the IP address has no longer the function of an end-point identifier. This significantly simplifies mobility since the node may easily change its HI and IP address bindings while moving. Furthermore, the security is improved since the name space is cryptographically based and it is thus possible to perform public key-based authentication using the HIs.

For further reading about the HIP protocol and HIP-related research projects, see Lundberg and Candolin (2003), Nikander, Ylitalo, and Wall (2003), and Helsinki University of Technology (2006).

Research Issues for Mobile Video Conferencing

Based on the analysis of current work in Featherstone and Zhang (2003), three major research issues for developing security and QoS provision in video conferencing systems for mobile computers were identified:

- **Mobility:** A satisfactory solution should provide a seamless QoS with mobility support and addressing of both wireless and wired networks.
- **Integrated Wired and Wireless Network:** Most mobile resource reservation schemes do not consider how QoS provision for wireless networks integrates with QoS provision on wired networks.
- **Security:** Systems providing both secure resource reservations as well as support for mobility are missing.

Testbed for Secure Mobile Multimedia

In Dutta et al. (2004b), a wireless Internet telephony and streaming multimedia testbed is set

up based on a testbed framework discussed in Dutta, Chen, Madhani, McAuley, NakaJima, and Schulzrinne (2001). Two types of mobility approaches are discussed: network-layer mobility by Mobile IP and application-layer mobility using SIP mobility. A moving mobile device is a target of potential attacks in different parts of the access and core network. Thus, a multi-layer security scheme is needed for providing comprehensive solutions that can possibly prevent any such attack and for supporting a dependable and secured multimedia application.

The testbed presented in Dutta et al. (2004b) is using an AAA protocol running on NAS (network access servers) and AAA servers to provide profile-based verification services. Additionally, a new protocol, PANA (protocol for carrying authentication for network access) (Yegin et al. 2005), is used for providing user-to-network access control. PANA provides access control mechanism to the mobile client and works as a front-end for the AAA server. It is implemented as a user level protocol to enable a flexible access control that works independently of any IPv4 or IPv6 layer 2 technology.

SIP-AAA Security Model

In the testbed two different security models were implemented. The first model is based on SIP and AAA and it was implemented in order to realize how SIP signaling can interact with an AAA infrastructure in a mobile environment. When the SIP server receives an SIP register message from the mobile device, it consults with the home AAA server for authentication and authorization. The profile verification database of the user is located in the home AAA server. In other words, in this model SIP registration is authenticated only after consulting with the AAA server. Normally, SIP registration is performed at the SIP server after the client has obtained a new address. The intention of the interaction between the SIP and the AAA server is to provide secure monitoring of user activities for accounting and auditing purposes.

IPSec-PANA-AAA Security Model

The second model, based on PANA and AAA, helps access control on the edge routers and interacts with IPSec to provide packet encryption. This model is a proposed multi-layered security scheme which provides packet-based encryption and application-layer authentication based on user NAI (network access identifier). In this model, PANA registration with the PANA server in the domain is performed when a user is roaming to a new network domain. In order to authenticate the user, the PANA server consults with the home AAA server either directly or indirectly through a local AAA server. As a result of successful PANA registration, an LSA (local security association) is established between the mobile device and the PANA server. Hereby, any further authentication required for intra-domain handoff is performed locally and quickly at the PANA server without the need to contact the home AAA server. The local authentication is also performed periodically with the intention to detect the event that the user disappears from the domain due to, for example, bad radio conditions.

The PANA server at the ERC (edge router & controller) maintains an association between the user identity, such as a NAI, and the lower-layer identity, such as an IP-address, for each user. The ERC has firewall functionality preventing packets sent from/to a mobile host not belonging to an authorized user to pass through the firewall. As a result of handoff, the association between the user identity and lower-layer identity dynamically changes. The ERC updates the access control list of the firewall if there is a change in the association and the resulting PANA registration or local authentication is successful. This prevents SIP register or SIP re-invite messages to pass through the firewall until the access control list is updated in the edge router.

In other words, PANA is used in the testbed to provide dynamic control of a router with firewall functionalities. Full network access is authorized only for hosts associated with authenticated PANA clients. An open firewall is closed immediately as a result of a failure of a periodical PANA re-authentication.

Packet Encryption and End-to-End Security

An IPSec-based encryption mechanism is used to secure data packets on the last hop in wireless networks. An IPSec tunnel is set up between the mobile client and the edge router. The PANA protocol is used for distributing IKE credentials to an authorized host. At first, the mobile client is authorized by PANA-based authentication. Then, the IKE credentials are carried in a PANA message and are transferred from the PANA authentication agent to the mobile client. These credentials are then used for setting up an IPSec tunnel between the mobile client and the access router. This provides a secure uni-cast communication channel in the access network including a wireless LAN segment. Since there is no need for a host to pre-configure the IKE credentials due to the dynamic distribution of the IKE credentials, mobile clients are enabled to smoothly roam among different domains.

In a mobile multimedia system it is essential to provide end-to-end security for both data and signaling. A combination of Mobile IP and IPSec solutions can be used to provide such an architecture, but such a combination is however not suitable for mobile systems since it will cause large overhead.

Real-time traffic is based on RTP/UDP. Thus, SRTP is used to provide encryption to different types of multimedia data. A secured RTP session requires the exchange of a separate RTP key between the mobile client and the correspondent host for a specific real-time media type. While registering with the SIP servers, SIP clients use PGP based authentication. The RTP key exchange is performed by an INVITE-exchange method using SIP signaling at the time of setting up calls. The RTP key is protected using a S/MIME mechanism. By these procedures, both data and signaling can be end-to-end secured without suffering from extra overhead.

CONCLUSIONS

Mobility and wireless communications introduce for multimedia applications similar problems with respect to security as for other applications. Advanced mobile terminals face new threats due to openness. Platforms are open for external software and content. Malicious software, like Trojan horses, viruses, and worms, has started to emerge. Fine-grained software authorisation has been proposed. Downloaded software may then access particular resources only through user authorisation. Vulnerabilities in OS implementation still remain a challenge because of difficulties in minimizing OS code running in privileged mode. Integrated hardware may be the solution.

Security aspects must be taken into account from the very beginning of the design of a mobile device for multimedia applications, not added on later. Security standards (SIP, secure RTP), technologies (IPSec) and testbed solutions for mobile multimedia are available, but significant improvements are still needed before acceptable security for mobile multimedia devices is achieved. Especially secure mobile video conferencing is still a research challenge.

REFERENCES

Baker, F., Lindell, B., & Talwar, M. (2000). *RSVP cryptographic authentication.* RFC 2747, IETF. Retrieved May 10, 2006, from http://www.ietf.org/rfc/rfc2747.txt

Barun, T., & Danzeisen, M. (2001). Secure mobile IP communication. In *Proceedings of 26th Annual IEEE Conference on Local Computer Networks (LCN'01)*, Tampa, FL, November 14-16 (pp. 586-593). Washington DC, USA: IEEE Computer Society.

Baugher, M., McGrew, S., Naslund, M., Carrara, E., & Norrman, K. (2004). *RTP: The secure real-time transport protocol (SRTP).* RFC 3711, IETF. Retrieved May 9, 2006, from http://www.ietf.org/rfc/rfc3711.txt

Braden, R., Zhang, L., Berson, S., Herzog, S., & Jamin, S. (1997). *Resource ReSerVation Protocol (RSVP).* RFC 2205, IETF. Retrieved May 10, 2006, from http://www.ietf.org/rfc/rfc2205.txt

Calhoun, P., Johansson, T., Perkins, C., Hiller,T., & McCann, Ed. P. (2005). *Diameter mobile IPv4 application*. IETF, RFC 4004. Retrieved May 8, 2006, from http://www.ietf.org/rfc/rfc4004.txt

Cremonini, M., Damiani, E., De Capitani di Vimercati, S., Samarati, P., Corallo, A., & Elia, G. (2005). Security, privacy, and trust in mobile systems and applications. In M. Pagani (Ed.), *Mobile and wireless systems beyond 3G: Managing new business opportunities* (pp. 312-340). Hershey, PA: IRM Press.

Dietze, C. (2005). The smart card in mobile communication: Enabler of next-generation (NG) services. In M. Pagani (Ed.), *Mobile and wireless systems beyond 3G: Managing new business opportunities* (pp. 221-253). Hershey, PA: IRM Press.

Dittmann, J., Wohlmacher, P., & Ackermann, R. (2001). Conditional and user specific access to services and resources using annotation watermarks. In R. Steinmetz, J. Dittman, & M. Steinebach (Eds.), *Communications and multimedia security issues of the new century* (pp. 137-142). Norwell, MA: Kluwer Academic Publishers.

Domenech, A. L. (2003). *Port-based authentication for wireless LAN access control.* Report of Graduation Degree. Department of Electrical Engineering Eindhoven University of Technology (TU/e), Eindhoven, Netherlands. Retrieved September 24, 2007, from http://people.spacelabs.nl/~alex/Port_Based_Authentication_for_Wireless_LAN_Access_Control.pdf

Duncan, M. V., Akhtari, M. S., & Bradford, P. G. (2004). Visual security for wireless handheld devices. In *JOSHUA Journal of Science & Health at The University of Alabama, 2*/2004. The University of Alabama.

Dutta, A., Chen, J., Madhani, S., McAuley, A., NakaJima, N., & Schulzrinne, H. (2001). Implementing a testbed for mobile multimedia. In *Proceedings of the IEEE Conference on Global Communications (GLOBECOM)*, San Antonio, Texas, November 25-29 (pp. 1944-1949). IEEE.

Dutta, A., Das, S., Li, P., McAuley, A., Ohba, Y., Baba, S., & Schulzrinne, H. (2004a). *Secured mobile multimedia communication for wireless Internet.* ICNSC 2004, Taipei.

Dutta, A., Zhang, T., Madhani, S., Taniuchi, K., Fujimoto, K., Katsube, Y., Ohba, Y., & Schulzrinne, H. (2004b). Secure universal mobility for wireless Internet. In *Proceedings of the 2nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH)*, Philadelphia, PA, October 1 (pp. 71-80). New York: ACM Press.

ETSI EN 302 304. (2004). *Digital video broadcasting (DVB): Transmission system for handheld terminals (DVB-H).* Retrieved September 26, 2007, from http://pda.etsi.org/pda/queryform.asp

Featherstone, I., & Zhang, N. (2003). Towards a secure videoconferencing system for mobile users. In the *5th European Personal Mobile Communications Conference*, Glasgow, Scotland, April 22-25 (pp. 477-481). London, UK: IEEE.

Gehrmann, C., Persson, J., & Smeet, B. (2004). *Bluetooth security.* Norwood, MA: Artech House, Inc.

Glass, S., Hiller, T., Jacobs, S., & Perkins, C. (2000). *Mobile IP authentication, authorization, and accounting requirements.* IETF, RFC 2977. Retrieved May 8, 2006, from http://www.ietf.org/rfc/rfc2977.txt

GSM Security Portal. (2005). Retrieved July 11, 2005, from http://www.gsm-security.net/

Havinga, P. J. M. (2000). *Mobile multimedia systems*. Ph.D. thesis, University of Twente, Netherlands, ISBN 90-365-1406-1.

Helsinki University of Technology. (2006). *InfraHIP project portal.* Retrieved May 10, 2006, from http://infrahip.hiit.fi

Hwu, J.-S., Chen, R.-J., & Lin, Y.-B. (2006). An efficient identity-based cryptosystem for end-to-end mobile security. *IEEE Transaction on Wireless Communications, 5*(9), 2586-2593.

Islam, R. (2005). *Enhanced security in mobile IP communication*. MSc Thesis, Department of Computer and Systems Sciences, Royal Institute of Technology, Stockholm, Sweden.

Jansen, W. A. (2003). Authenticating users on handheld devices. In the *15th Annual Canadian Information Technology Security Symposium (CITSS),* Ottawa, Canada, May 12-15. Ottawa, Canada: Communications Security Establishment. Retrieved September 24, 2007, from http://csrc.nist.gov/mobilesecurity/publications.html#MD

Kaspersky Lab Products & Services. (2007). *Kaspersky anti-virus mobile*. Retrieved September 24, 2007, from http://www.kaspersky.com/antivirus_mobile

Kharif, O. (2005). May I see your voice, please? In *BusinessWeek online*, April 20. Retrieved July 11, from http://www.businessweek.com/technology/content/apr2005/tc20050420_1036_tc024.htm?campaign_id=rss_techn

Kivinen, T., & Tschofenig, H. (2006). *Design of the IKEv2 Mobility and Multihoming (MOBIKE) Protocol, draft-ietf-mobike-design-08.txt.* RFC 4621, IETF. Retrieved March 3, 2006, from http://www.ietf.org/internet-drafts/draft-ietf-mobike-design-08.txt

Lee, B.-G., Choi, D.-H., Kim, H.-G., Sohn, S.-W., & Park, K.-H. (2003). Mobile IP and WLAN with AAA authentication protocol using identity-based cryptography. *10th International Conference on Telecommunications ICT 2003, 1*, Colmar, France, February 23-March 1 (pp. 597-603). USA: IEEE.

Liu, Q., Reihaneh, S.-N., & Sheppard, N. P. (2003). Digital rights management for content distribution. In *Proceedings of the Australasian Information Security Workshop Conference on ACSW Frontiers 2003, 21*, Adelaide, Australia, February 1 (pp. 49-58). Darlinhurst, Australia: Australian Computer Society, Inc.

Lundberg, J., & Candolin, C. (2003). Mobility in the host identity protocol (HIP). In *Proceedings of the International Symposium on Telecommunications (IST2003),* Isfahan, Iran, August (pp. 754-757). Iran: Iran Telecom Research Center.

Meyer, U., & Wetzel, S. (2004). On the impact of GSM encryption and man-in-the-middle attacks on the security of interoperating GSM/UMTS networks. In the *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2004), 4*, Barcelona, Spain, September 5-8 (pp. 2876-2883). USA: IEEE. Retrieved July 12, 2005, from http://www.cdc.informatik.tu-darmstadt.de/~umeyer/UliP-IMRC04.pdf

MITA. (2002). *Mobile Internet technical architecture: The complete package*. Nokia, Finland, ISBN 951-826-669-7.

Mobile Broadcast/Multicast Service (MBMS). (2004, August). White Paper, TeliaSonera. Retrieved July 11, 2005, from http://www.medialab.sonera.fi/workspace/MBMSWhitePaper.pdf

MobileCloak™ Web Portal. (2005). Retrieved July 11, 2005, from http://www.mobilecloak.com

Moskowitz, R., & Nikander, P. (2006). *Host Identity Protocol (HIP) Architecture*. RFC 4423, IETF.

Muller, T. (1999). *Bluetooth security architecture: Version 1.0*. Bluetooth White Paper, Document # 1.C.116/1.0. Retrieved September 26, 2007, from http://www.bluetooth.com/NR/rdonlyres/C222A81E-D9F9-48CA-91DE-9C81F5C8B94F/0/Security_Architecture.pdf

Nasir, A., & Mah-Rukh, M.-R. (2006). Internet mobility using SIP and MIP. *Proc. Third International Conference on Information Technology: New Generations ITNG'06*, Las Vegas, NV, April 10-12 (pp. 334-339). USA: IEEE Computer Society.

Nikander, P. (2001). *Denial-of-service, address ownership, and early authentication in the IPv6 world*. Retrieved May 10, 2006, from http://www.tml.tkk.fi/~pnr/publications/cam2001.pdf

Nikander, P., Ylitalo, J., & Wall, J. (2003). Integrating security, mobility, and multi-homing in a HIP way. *Proceedings of Network and Distributed Systems Security Symposium (NDSS'03),* San Diego, USA, February (pp. 87-99). Reston, USA: Internet Society.

Olzak, T. (2005). *Wireless handheld device security*. Retrieved July 11, 2003, from http://www.securitydocs.com/pdf/3188.PDF

Ong, C. S., Nahrstedt, K., & Yuan, W. (2003). Quality of protection for mobile multimedia applications. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME2003)*, Baltimore, MD, July 7-9 (pp. 137-140). IEEE. Retrieved July 11, 2005, from http://cairo.cs.uiuc.edu/publications/papers/ICME03-chui.pdf

Perelson, S., & Botha, R. (2004). An investigation into access control for mobile devices. In H. S. Venter, J. H. P. Eloff, L. Labuschagne, & M.M. Eloff (Eds.), *ISSA 2004 Enabling Tomorrow Conference, Peer-Reviewed Proceedings of the ISSA 2004 Enabling Tomorrow Conference, Information Security South Africa (ISSA),* Gallagher Estate, Johannesburg, South Africa, June 30-July 2. CDROM, ISBN 1-86854-522-9. ISSA. Retrieved July 11, 2005, from http://www.nmmu.ac.za/rbotha/Pubs/docs/LC_017.pdf

Perkins, C. (2002). *IP mobility support for IPv4, IETF*. RFC 3344. Retrieved May 8, 2006, from http://www.ietf.org/rfc/rfc3344.txt

Phillips, A., Nelson, B., Enfinger, F., & Steuart, C. (2005). *Guide to computer forensics and investigations* (2nd ed.). USA: Course Technology Press.

Pulkkis, G., Grahn, K., Karlsson, J., Martikainen, M., & Daniel, D. E. (2005). Recent developments in WLAN security. In M. Pagani (Ed.), *Mobile and wireless systems beyond 3G: Managing new business opportunities* (pp. 254-311). Hershey, PA: IRM Press

Robbins, J. (2005). *An explanation of computer forensics*. Retrieved July 11, 2005, from http://www.computerforensics.net/forensics.htm

Rosenberg, J., Schulzrinne, H., Camarillo, G, Johnston, A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). *SIP: Session initation protocol*. RFC 3261, IETF. Retrieved May 9, 2006 from http://www.ietf.org/rfc/rfc3261.txt

Salsano, S., Veltri, L., & Papalilo, D. (2002). SIP security issues: The SIP authentication procedure and its processing load. *IEEE Network, 16*(6), 38-44.

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). *RTP: A transport protocol for real-time applications*. RFC 3550, IETF. Retrieved May 8, 2006, from http://www.ietf.org/rfc/rfc3550.txt

Setec Portal. (2005). Retrieved July 11, 2005, from http://www.setec.fi

Stamp, M. (2003). Digital rights management: The technology behind the hype. Journal of Electronic Commerce Research, 4(3), 202-212. Long Beach, CA: California State University Long Beach. Retrieved May 12, 2006, from http://home.earthlink.net/~mstamp1/DRMpaper.pdf

Sun, J., Howie, D., Koivisto, A., & Sauvola, J. (2001a). A hierarchical framework model of mobile security. *Proc. 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communication, 1,* San Diego, CA, September 30-October 3 (pp. 56-60). USA: IEEE.

Sun, J., Howie, D., Koivisto, A., & Sauvola, J. (2001b). Design, implementation, and evaluation of Bluetooth security. *Proc. IEEE International Conference on Wireless LANs and Home Networks,* Singapore, December 5-7 (pp. 121-130). USA: IEEE. Retrieved July 11, 2005, from http://www.mediateam.oulu.fi/publications/pdf/87.pdf

Sundaresan, H. (2003). *OMAP™ platform security features.* White Paper. Retrieved July 1, 2005, from http://focus.ti.com/pdfs/wtbu/omapplatformsecuritywp.pdf

Symantec. (2003). *Wireless handheld and smartphone security.* Retrieved September 26, 2007, from http://whitepapers.zdnet.co.uk/0,1000000651,260085794p,00.htm?r=1

Symantec Corporation. (2005). *Symantec AntiVirus™ for handhelds annual service edition, Symantec antiVirus for handhelds safeguards Palm and Pocket PC mobile users.* Retrieved July 9, 2005, from http://www.symantec.com/sav/handhelds/ and http://www.symantec.com/press/2003/n030825.html

Talwar, V., & Nahrstedt, K. (2000). Securing RSVP for multimedia applications. *Multimedia Security Workshop ACM Multimedia,* Los Angeles, CA, October 30-November 3 (pp. 153-156). USA: ACM Press.

Taylor, L. (2004-2005). *Handheld security, part I-V.* Retrieved September 25, 2007, from http://www.pdastreet.com/articles/2004/12/2004-12-6-Handheld-Security-Part.html

Trusted Computing Group. (2004). *Security in mobile phones.* Retrieved July 1, 2005, from http://www.trustedcomputinggroup.org/downloads/whitepapers/TCG-SP-mobile-sec_final_10-14-03_V2.pdf

WEPCrack. (2007). Retrieved September 26, 2007, from http://wepcrack.sourceforge.net

Wi-Fi Alliance Portal. (2005). Retrieved July 11, 2005, from http://www.wi-fi.org

Wikipedia. (2005). *Secure sockets layer (SSL) and transport layer security (TLS).* Retrieved July 10, 2005, from http://en.wikipedia.org/wiki/Secure_Sockets_Layer

Wikipedia, the free encyclopedia. Retrieved July 11, 2005, from http://www.wikipedia.org

Wikipedia. *Computer forensics.* Retrieved July 11, 2005, from http://en.wikipedia.org/wiki/Computer_forensics

Wu, T., Wu, S., Huang, H., & Gong, F. (1999). Securing QoS: Threats to RSVP messages and their countermeasures. *Proceedings of IWQoS'99 – Seventh International Workshop on Quality of Service,* UCL, London, May 31-June 4 (pp. 62-64). USA: IEEE.

Xydis, T. G., & Blake-Wilson, S. (2002). *Security comparision: Bluetooth™ Communications vs. 802.11.* Retrieved July 11, 2005, from http://ccss.isi.edu/papers/xydis_bluetooth.pdf

Yegin, A., Ohba, Y., Penno, R., Tsirtsis, G., & Wang, C. (2005). *Protocol for carrying authentication for network access (PANA) requirements.* RFC 4058, IETF.

Zeng, W., Zhuang, X., & Lan, J. (2004). Network friendly media security: Rationales, solutions, and open issues. *2004 International Conference on*

Image Processing ICIP '04, 1, Singapore, October 24-27 (pp. 565-568). USA: IEEE.

Zhang, L., Lambertsen, G., & Yamada, T. (2003). Security scenarios within IP-based mobile multimedia metropolitan area network. In *Proceedings of Global Telecommunications Conference, GLOBECOM '03, 3,* San Francisco, CA, December 1-5 (pp. 1522-1526). Piscataway, NJ: IEEE Operations Center.

APPENDIX: AVAILABLE SECURITY SOLUTIONS

Market offers several products as a solution to the increasing security problems of mobile devices. These products are designed to solve individual or more comprehensive security problems like unauthorized access to data or device, viruses, and unencrypted data transfers. The following list of the product groups reveal versatility of the mobile device security solutions (Douglas, 2004; Taylor, 2004, part III):

- Platform security products
- Authentication products
- Encryption products
- Anti-virus products
- Bit wiping products
- Firewall products
- VPN products
- Wireless security products
- Forensic products
- Database security products
- Centralized security management products
- Backup/restore products

In this section, available software products for platform security, authentication, encryption, anti-virus, VPNs (virtual private networks), firewall, and computer forensics are presented. The specific features of some products from each category are briefly described. In addition, characteristics and benefits of three well-known multifunctional security products for mobile devices, Pointsec, Bluefire and PDASecure, will be handled. Product information was searched during spring 2005.

Platform Security Solutions

There are different embedded on-chip security solutions, but mostly the security solution relies on a combination of hardware and software components. In the following, the two new security solutions, Texas Instruments OMAPTM Platform (Sundaresan, 2003) and Intel Wireless Trusted Platform (Intel, 2004), are shortly described.

The TI solution relies on three layers of security: application layer security, operating system layer security, and on-chip hardware security. The platform includes four main security features:

- **Secure Environment:** This feature enables secure execution of critical code and data. Sensitive information is hidden from the outside world with the use of four security components. These are *secure mode*, *secure keys*, *secure ROM*, and *secure RAM*. Secure mode allows secure execution of "trusted" code via on-chip authentication. Secure mode can be seen as a 3rd privilege level. Secure keys, ROM, and RAM are accessible only in secure mode. The keys are OEM specific one-time programmable keys used for authentication and encryption. Secure ROM is a storage feature using secure keys, key management and cryptographic libraries. Secure RAM is used for running OEM specific authenticated code.
- **Secure Boot/Flash:** The secure boot/flash process prevents security attacks during device flashing and booting. The authentication process must guarantee the origin and the integrity of the software stored on the platform. The process also prevents execu-

tion of any spoofed software code.

- **Run-Time Security:** The secure environment has been integrated into the OS. Thus, OS applications for short security critical tasks can be performed. Such tasks include encryption/decryption, authentication, and secure data management.
- **Crypto Engine:** Hardware crypto engines are used to enhance performance and security. Available crypto engines are DES/3DES, SHA1/MD5, and RNG. Two configuration modes are available: secure mode for short and/or high security level applications, and user mode for long and/or low-level security applications.

The building blocks of the Intel platform consists of performance primitives (hardware) and cryptographic primitives (optimized software). The solution enables services such as trusted boot and safe processing of secrets, protected key storage, and attestation measuring the security status of the platform during trusted boot.

The components of the platform include:

- **Trusted Boot ROM:** Validates the integrity of the platform and boots it into the right configuration.
- **Wireless Trusted Module:** This is the module where the secrets are processed. This suite of cryptographic engines include random number generation, symmetric and asymmetric cryptography, key creation, key exchange, digital signature operations, hashing, binding, and a monotonic counter.
- **Security Software:** A security software stack enables the OS and applications to access the platform resources through standard cryptographic APIs.
- **Protected Storage:** Protected storage in system flash allows secure non-volatile storage of secrets.
- **Physical Protection:** Physical protection is supported by integration of the security hardware in a single device and by packaging the discrete components into a single physical package.

Authentication Products

Unauthorized access has not always been recognized as a security risk especially among private users, even though mobile devices such as PDAs and smartphones are small, portable, and easily lost or stolen. These features lead to a high risk of vital data loss. From this point of view, it is easy to understand the necessity of authentication.

There are several authentication methods (Douglas, 2004, p.13):

- Electronic signature authentication
- Picture-based password authentication
- Fingerprint authentication
- Card-based authentication
- Storage-card-based certificate authentication
- Legacy host access

This section concentrates on products for electronic signature, picture-based password, and fingerprint authentication. A summary of authentication product examples for Pocket PCs and smartphones is presented in Tables A1-A3.

Electronic Signature Authentication Products

Electronic signature authentication has several benefits. It provides a high level of security and an electronic signature is, from the user's point of view, a simple password which cannot be forgotten. The main problem with this kind of technology is that the biographic signature is varying from time to time, which causes the possibility of access denial.

To mention a couple of these kinds of solutions, Romsey Associates Ltd. offers PDALok with dynamic signature recognition and Transaction

Security Inc. has PDA Protect with crypto-sign pattern recognition.

When logging on or synchronizing of data, PDALok forces users to set PIN (personal identification number) or signature. PDALok's biometric technology, dynamic signature recognition, monitors different characteristics of writing as rhythm, speed, pressure, flow, and acceleration. The software is compatible with almost every mobile device running on Window CE and Palm OS (Romsey Associates Ltd., 2005).

Topaz Systems Inc. (2005) provides a more traditional signature authentication alternative. The set with separate writing pads includes software and inkless pen such as SignatureGem. Features of signature authentication products are shown in Table A1.

Picture-Based Password Authentication

Instead of the traditional alpha-numeric login passwords and PINs, Pointsec's PicturePIN and sfr Gesellschaft für Datenverarbeitung mbH's visKey Palm OS picture-based authentication are alternative solutions. Picture-based login style is easier to remember and more user-friendly for mobile devices with small screens and mini keyboards.

Pointsec allows users to select a password consisting of a combination of icons (Figure A1). The position of the Pointsec's icons changes place each time the device is activated (Pointsec Mobile Technologies AB., 2005).

Alternatively, visKey Palm OS users select several desired spots in an image (Figure A2). Before creating the password, visKey Palm OS splits up the selected image into cells. Each cell represents a single character (sfr Gesellschaft für Datenverarbeitung mbH, 2005). The user identifies to the device by choosing the same order of clicks. Features of picture-based password authentication products are shown in Table A2.

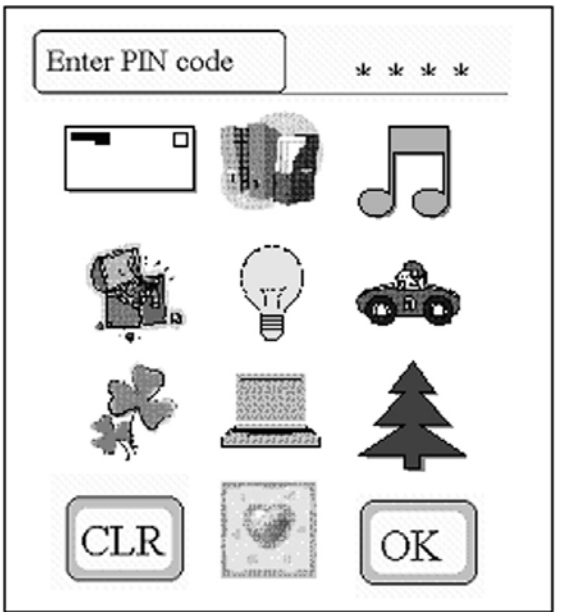
Fingerprint Authentication

Fingerprint-based identification is the oldest method of biometric techniques. The uniqueness of a human fingerprint prevents effectively

Table A1. Signature authentication products

Company	Product Name	Feature / Function
Certicom Corporation	Security Builder	Cross-platform cryptographic module, which allow developers to handle all encryption, decryption, digital signatures and message authentication codes.
Communication Intelligence Corporation	InkTools	Developers' kit with electronic ink capture and display, signature verification, encryption, ink compression and other algorithms.
Romsey Associates Ltd.	PDALok	Uses biometric technology, Dynamic Signature Recognition to prevent unauthorized access to Pocket PC.
Transaction Security Inc.	PDA Protect	Crypto-Sign™ is a biometric pattern recognition technology based upon the submission of a secret sign.
VASCO	Digipass	Digipass offers strong user authentication and electronic signatures.

Figure A1. Pointsec's PicturePIN picture-based authentication



forgery attempts. The security level of fingerprint authentication is depending on such factors as the quality of scanning and the visual image recognition (ROSISTEM, 2005).

Veridt LLC produces fingerprint authentication solutions, for example BioHub for Pocket PC and BioSentry for Compaq iPAQ Pocket PC (Veridt, LLC., 2005). BioHub is a combination of hardware and software for fingerprints authentication. The product contains a plug-in capture device for fingerprint imaging and software which identifies and stores fingerprint patterns.

BioSentry is a separate device into which a Pocket PC is placed. The device compares the fingerprint to the fingerprint minutiaes saved in the Pocket PC's memory. The minutiaes of a fingerprint are defined from a scanned image and the coordinates of these minutiaes are calculated. Finally these coordinates are encrypted and saved. The device uses an encryption algorithm which meets the requirements of AES-128, AES-192, and AES-256 standards. Features of fingerprint authentication products are shown in Table A3.

Encryption Products

One of the simplest methods to protect sensitive data is encryption. Pointsec for Pocket PC provides a solution for real-time encryption of data on both PDAs and different types of memory cards without any user interaction. Data can only be accessed or decrypted with proper authentication (Pointsec Mobile Technologies AB, 2005).

Trust Digital 2005 offers the possibility of full background synchronization of encrypted data (Trust Digital, 2005). Airscanner Mobile Encrypter secures data with optional password methods. Instead of the traditional individual file and folder encryption, where each file can be encrypted/decrypted with its own password, one global password can be applied. After sign-on, transparent and automatic volume encryption takes place within a user-defined time (Airscanner® Corp., 2005). Features of storage encryption products are shown in Table A4.

Anti-Virus Products

Major anti-virus software companies started to develop anti-virus products since the first Symbian OS mobile phone virus, the Cabir worm, was found in 2004 (Kaspersky Lab, 2005).

F-Secure manufactures anti-virus security software, especially for Nokia's mobile phones, for example, Series 60/90 and Nokia 9300/9500 Communicator Series. Symantec offers anti-virus products for Palm OS- and Microsoft Pocket PC-compatible devices. To prevent infection, F-Secure Mobile Anti-Virus and Symantec AntiVirus for Handhelds scan all files automatically and transparently during modification and transference of data, without user intervention. When an infected file is detected, the file is immediately quarantined by the system to protect all other data.

F-Secure Mobile Anti-Virus for Nokia 9300/9500 Communicator Series provides automatic updates for the virus signature database over a secure HTTPS data connection or incrementally with SMS messages.

Figure A2. Gesellschaft für Datenverarbeitung mbH's visKey Palm OS picture-based authentication

Figure A2a

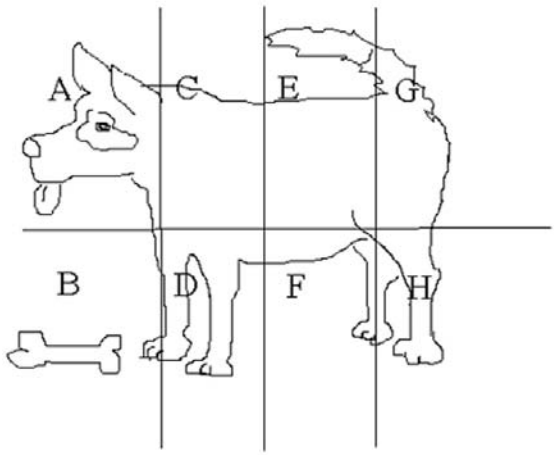


Figure A2b

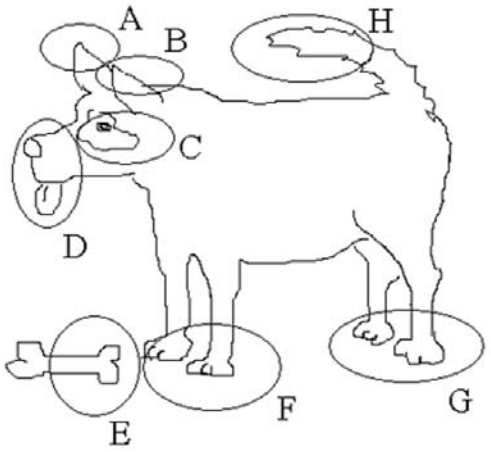


Table A2. Picture-based password authentication

Company	Product Name	Feature / Function
Pointsec Mobile Technologies	Pointsec for Smartphones, Pointsec for Pocket PC, Pointsec for Palm OS, Pointsec for Symbian OS,	The picture-based authentication software enables users to select a password consisting of a combination of icons.
sfr Gesellschaft für Datenverarbeitung mbH	visKey Palm OS	The picture-based authentication software enables users to select several desired spot in an image.
Softava	PicturePassword	User authentication with one click on the desired point of an image.

Table A3. Fingerprint authentication products

Company	Product Name	Feature / Function
Hewlett-Packard Development Company, L.P.	HP iPAQ 5450, 5455 and 5555 Pocket PC.	The HP iPAQ series Pocket PC with inbuilt biometric fingerprint reader.
Veridt, LLC	BioHub	BioHub is a combination of hardware and software for fingerprints authentication solutions.
Veridt, LLC	BioSentry	A fingerprint verification "jacket" for Pocket PC.

Table A4. Storage encryption products

Company	Product Name	Feature / Function
Airscanner Corp.	Airscanner Mobile Encrypter	File and folder encryption for Pocket PC
Asynchrony	PDA Defense	Encrypts all databases, files and memory cards on Pocket PC.
Certicom Corporation	movianCrypt	Encrypts and locks all information without impeding performance or usability on Pocket PC.
Glück & Kanja Group	CryptoEx Pocket	Secure email communication and file storage encryption on Pocket PC.
PC Guardian Technologies, Inc.	PDASecure	Provides data encryption and access control. The software encrypts data based on administrator and user preferences on Pocket PC, Palm OS, smartphone, Symbian and BlackBerry.
Pointsec Mobile Technologies	Pointsec for Pocket PC, Pointsec for Smartphone, Pointsec for Symbian OS	Provides real-time encryption of all data on both their PDAs and all removable media on Pocket PC, smart phone, Symbian OS.
SoftWinter	Sentry 2020 for Pocket PC	Provides encrypted virtual volumes. Data stored on a Sentry volume is transparently encrypted/decrypted.
Cranite Systems, Inc.	WirelessWall	provides FIPS 140-2 certified AES data encryption for Pocket PCs.

With LiveUpdate Wireless feature from Symantec AntiVirus for Handhelds, users are enabled to download virus definitions and Symantec product updates directly to their mobile device with a wireless Internet connection (F-Secure Corporation, 2005; Symantec Corporation, 2005).

Instead of commercial products, BitDefender offers free anti-virus software, BitDefender Antivirus—Free Edition for Windows CE and Palm OS. They are freeware, which means that no license is required for using the products. Features of anti-virus products are shown in Table A5 (BitDefender, 2007).

VPN (Virtual Private Network) and Firewall Products

Wireless networks cause a remarkable security risk because the transmitted data over the air can be easily exploited by outsiders. Secure VPNs use cryptographic tunneling protocols to ensure sender authentication, confidentiality, and integrity of data.

Columbitech's Wireless VPN is a client/server software architecture in which a secure encrypted virtual tunnel between the VPN client (mobile device) and the VPN Server is created to secure the data traffic between the devices. Authentication,

validation, and the transmitted data is encrypted with WTLS (Wireless Transport Layer Security). WTLS is applied because it has better performance than IPSec (IP Security) over wireless networks. Wireless VPN uses the AES algorithm, with up to 256 bit encryption keys.

The software uses PKI (public key infrastructure), smart cards, biometric identification, and one-time passwords as authentication mechanisms. In order to prevent unauthorized access Columbitech's Wireless VPN has firewall functionality on both client and server. The client integrity is ensured with security status monitoring before the VPN tunnel is created. See Figure A3 (Columbitech AB, 2005).

Compared to the more common VPN, that uses IPSec technology, the newer VPN with SSL (secure sockets layer) cryptographic protocol is easier for administrators and users to set up and manage. The benefit of using SSL VPN instead of IPSec VPN is that SSL does not require any client software to be installed on each remote device because SSL is widely supported on most Web browsers (Ferraro, 2003).

Intoto's iGateway SSL-VPN allows users to securely access to key corporate resources at remote locations and create a secure encrypted virtual tunnel through any standard Web browser. iGateway SSL-VPN offers alternative authentication methods: RADIUS (Remote Authentication Dial-In User Service), LDAP (Lightweight Directory Access Protocol), Active Directory, Windows NTLN (NT LAN Manager), and digital certificates (Intoto Inc., 2005).

There are also Linux-based firewalls for PDAs on the market. Firewall programs, such as iptables or netfilter, are included on Familiar's Linux distribution list. Features of VPN and firewall products are shown in Table A6 (The Familiar Project, 2005; Villamil, 2005).

Forensic Analysis Products

While a large variety of forensic analysis software is available for personal computers, the range of solutions is much more limited for mobile devices. The problem is not only fewer software products for PDAs. These products operate only in most common families of PDAs.

The forensic analysis products have three main functionalities: acquisition, examination, and reporting. Only few products have all these mentioned functionalities. In such cases, several software products have to be purchased in order to accomplish the forensic examination process. The forensic analysis products need full access to the devices, in order to start acquisition of data. If the examined device is protected with

Figure A3. Columbitech's Wireless VPN solution

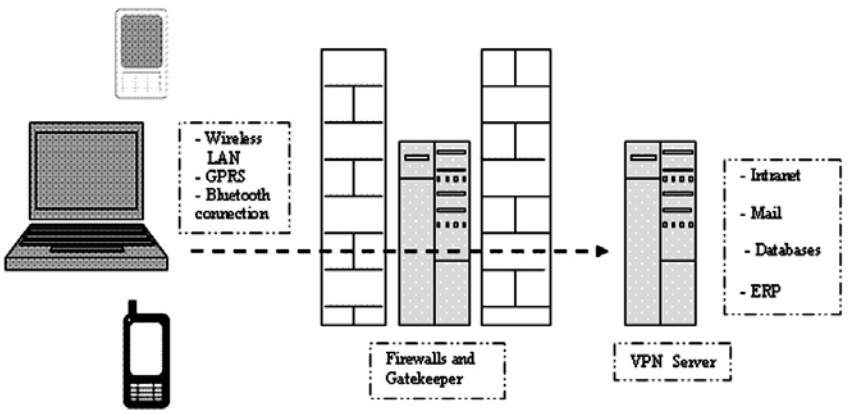


Table A6. VPN and firewall products

Company	Product Name	Feature / Function
Certicom Corporation	MovianVPN	The client for Palm and Pocket PCs operates with VPN gateways from Check Point, Cisco, and Nortel.
Check Point Software Technologies Ltd.	VPN-1 Secure Client	Allow users to securely access resources protected by VPN-1 gateways and provides a personal firewall on Pocket PCs.
Columbitech	Wireless VPN	A client/server software architecture, which create a secure WTLS encrypted virtual tunnel between the VPN client and the VPN Server, to secure the data traffic between the devices.
Ecutel Inc.	Viatores Mobile IP VPN	A client-server VPN solution. The software is based on Mobile IP and IPSec.
Funk Software Inc.	AdmitOne VPN Client for Pocket PC	offers advanced IPsec and IKE technologies to secure the transmission of data between Pocket PCs and the VPN server.
Intoto	iGateway SSL-VPN	enable users to create a secure encrypted virtual tunnel through at any standard web browser. The software offers alternative authentication methods: RADIUS, LDAP, Active Directory, Windows NTLN and digital certificates.
Symbol Technologies Inc.	AirBEAM Safe	A wireless VPN solution for full "end-to-end" security for Pocket PCs. It delivers strong, end-to-end encryption between the corporate firewall and the application server, using industry-standard authentication methodology.
the Familiar Project	Familiar v0.8.2	The firewall program "iptables" is included on Familiar's Linux distribution. Familiar v0.82 supports the iPAQ h series, Siemens Simpad and Sharp Zauri.

Table A7. Forensic analysis products

Company	Product name	Feature
Paraben Corporation	PDA SEIZURE version 3.0.2.43	Forensic acquisition, examination and reporting of data for Palm OS and Pocket PC.
Guidance Software Inc.	ENCASE Version 5	Forensic acquisition, examination and reporting of data for Palm OS.
Symantec Corporation	PDD	Memory imaging and forensic acquisition for Palm OS.
PalmSource Inc.	Palm OS Emulator (POSE)	Forensic examination and reporting for Palm OS.
Compelson Laboratories	MOBILedit!	Forensic gathering and reporting of data for mobile phones.

some authentication method, cracking software is needed.

Paraben's PDA Seizure and Guidance Software Inc's EnCase Forensic Version 5 are forensic analysis tools that allow forensic examiners to acquire, examine, and analyze data. The PDA Seizure tool supports Palm OS, Windows-based OS, and BlackBerry devices, while Encase Forensic currently only supports Palm OS. Both of these mentioned tools have possibility for physical and logical acquisition of data. Physical acquisition means complete bit-by-bit copying of physical storage, for example a disk drive. Logical acquisition also means exact copying of logical storage objects, such as files and folders (Ayers & Jansen, 2004, p.14).

The EnCase Evidence file, created by EnCase, is suitable for the computer security industry and even for the law enforcement. EnCase allows users to determine how or what kind of data will be presented. This depends on the examination purposes. More Forensic analysis products are shown in Table A7 (Guidance Software, Inc., 2005; Paraben Corporation, 2007).

Multi-Functional Products

Multi-functional products are developed to solve comprehensively all security needs of mobile devices. From an administrator's point of view, these kind of products are appealing while there can be saved a lot of resources in terms of effective central administration. In this part of the chapter, three well-known commercial multi-functional solutions from Pointsec Mobile Technologies, PC Guardian Technologies, and Bluefire Security technologies will be presented.

Pointsec has solutions for the following devices and purposes: smartphone, Palm OS, Symbian OS, Pocket PC, and media encryption. PC Guardian Technologies software PDASecure is applicable to the smartphones, Pocket PCs, Palm OS, BlackBerry, and Symbian OS. Bluefire Security Technologies software Bluefire Security Suite 3.5 operates on Palm OS and Windows-based mobile platforms.

One of the most important features of all these multi-functional solutions is the third party management possibility. Administrators can enforce, update, and deploy the company security policy, which is applied also on desktops and laptops. Wished security level is ensured when a user

does not adjust the policy according to their preferences. This means that the security system is user transparent and does not contain or demand any user intervention. Pointsec's software is also compatible with all major third-party management tools, which means that administrators can easily integrate their Pointsec's software into their existing system.

Available third-party management system is used to give remote help such as password reset for users. Bluefire highlights their real-time logging function, which enables the administrator to log and monitor all suspicious activities such as password attempts and changes in firewall security level.

The key function of PDASecure, Bluefire, and Pointsec software is encryption. This feature ensures high security level because all the data can be automatically and immediately encrypted before they are stored or transferred and decrypted automatically by an authenticated user. All of the three example products use the AES (Advanced Encryption Standard) algorithm, the U.S. government approved cryptographic standard with strong 128-bit encryption keys. PDASecure even allows users or administrators to choose from six alternative encryption algorithms.

Authenticated synchronization has been mentioned as an essential part of PDASecure's and Pointsec's mobile device software. Pointsec's ActiveSync system requires the users to authenticate themselves to the mobile device before synchronization with a PC can be started.

None of the example software emphasizes the anti-virus or VPN part of their solution. In this case only Bluefire offers anti-virus and VPN software as an additional feature (Bluefire Security Technologies, 2005; PC GuardianTechnologies, 2005; Check Point Software Technologies LTD., 2007). In Table A8, features of Pointsec, PDASecure, and Bluefire Security Suite 3.5 are shown.

Case Study: Encryption Software Performance Measurements

This case study presents performance measurements on the security software Pointsec for Symbian OS. The purpose was to measure the influence of Pointsec on data communication performance of Symbian OS. Pointsec is presented in more detail in the section, *Add-on Security Software*. According to Pointsec Mobile Technologies, the Pointsec security software should not reduce speed or other performance measures even when the strong 128-bit AES encryption is used to protect the information in the device and in memory cards.

Measurements

However, security solutions may reduce data communication performance measures of mobile operating systems, such as download speed and connection times. These performance measures were measured for a Pointsec security software installation in a Nokia Communicator 9500 for:

- downloading a 4.92 MB file;
- connection to an e-mail server (penti.arcada.fi) with imaps based e-mail client software;
- connection to a www site (www.nokia.com);
- connection to a ssh server (penti.arcada.fi) with a putty ssh client.

All four performance measures were measured six times with and without installed Pointsec security software for two different access network types, WLAN and GPRS. The network bandwidths were 11 Mbit/s for the WLAN and 56 kbit/s for GPRS. Measurement results are presented in Tables A9-A12.

Table A8. Multi-functional products

Features	Company: Pointsec Product: Pointsec for Pocket PC, Smartphone, Symbian OS and PalmOS	Company: PC Guardian Technologies Product: PDASecure	Company: Bluefire Security Technologies Product: Bluefire Mobile Security Suite 3,5
Security policy enforcement from one central location	X	X	X
Remote Help and 3rd party distribution	X	X	X
FIPS 140-2 certified, AES algorithm with 128 bit encryption key	X	X	X
Automatic and immediate encryption	X	X	X
Memory card encryption	X	X	X
Picture-based passwords authentication	X		
Alphanumeric password authentication	X	X	X
Authenticated activeSync	X	X	
Antivirus software / virus protection		X	X
Firewall			X
VPN			X

Usefulness of Measurement Results

Conditions cannot be assumed to be equal for different measurements, since the download speed and connection times were measured for data communication through the public Internet. The utilization of Internet during a measurement session is not deterministic. Measurement results have been considered to be useful if standard deviation is less than 10% of the calculated average for measurements with the same mobile device configuration. Standard deviation exceeded 10% of calculated average only in one measurement case, GPRS connection to an e-mail server without Pointsec security software installed, being about 15% of calculated average (see Table A10).

Measured Degradation of Data

Communication Performance Caused by Pointsec

The influence of Pointsec was considered to be noticeable if the intervals defined by measured average and standard deviation do not overlap with and without Pointsec for otherwise the same mobile device configuration. Noticeable performance degradation was measured only for connection time to a Web site, about twice as long as for a GPRS connection and about 17% longer for a WLAN connection (see Table A11). However, the influence of:

- the traffic load on the Internet, and
- the load on the selected Web server

during carried out performance measurements is unfortunately unknown.

The measurements can thus be considered to support the view of the provider of Pointsec security software, that the performance degradation from this security software is insignificant on a Symbian device.

REFERENCES

Airscanner® Corp. (2005). *Airscanner® Mobile Encrypter V2.2b (BETA) user's manual.* Retrieved July 8, 2005, from http://airscanner.com/downloads/encrypter/EncrypterManual.htm

Ayers, R., & Jansen, W. (2004). *PDA forensic tools: An overview and analysis.* Retrieved July 6, 2005, from http://csrc.nist.gov/publications/nistir/nistir-7100-PDAForensics.pdf

BitDefender. (2007). *BitDefender free edition for Windows CE—AntiVirus Freeware, BitDefender Free Edition for PALM OS—AntiVirus Freeware.* Retrieved September 27, 2007, from http://www.johannrain-softwareentwicklung.de/bitdefender_antivirus_free_edition_pour_windows_ce.htm or http://www.johannrain-softwareentwicklung.de/bitdefender_antivirus_free_edition_pour_palm_os.htm

Bluefire Security Technologies. (2005). *Wireless device security products.* Retrieved July 9, 2005, from http://www.bluefiresecurity.com/products.html

Check Point Software Technologies LTD. (2007). *Welcome Pointsec and Reflex Magnetics customers!* Retrieved September 27, 2007, from http://www.checkpoint.com/pointsec/

Columbitech AB. (2005). *Columbitech wireless VPN, technical description.* Retrieved September 26, 2007, from http://www.columbitech.com/img/2007/5/7/5300.pdf

Douglas, D. (2004). *Windows mobile-based devices and security: Protecting sensitive business information.* Retrieved July 1, 2005, from http://download.microsoft.com/download/4/7/c/47c9d8ec-94d4-472b-887d-4a9ccf194160/6.%20WM_Security_Final_print.pdf

Ferraro, C. I. (2003). *Choosing between IPsec and SSL VPNs.* Retrieved July 10, 2005, from http://searchsecurity.techtarget.com/qna/0,289202,sid14_gci940324,00.html

Guidance Software, Inc. (2005). *EnCase Forensic Version 5.* Retrieved July 8, 2005, from http://www.guidancesoftware.com/products/ef_index.asp

Intel. (2004). *Intel wireless trusted platform: Security for mobile devices.* White Paper. Retrieved September 26, 2007, from http://whitepapers.zdnet.co.uk/0,1000000651,260091578p,00.htm

Intoto Inc. (2005). *iGateway SSL-VPN.* Retrieved July 9, 2005, from http://www.intoto.com/product_briefs/iGateway%20SSL%20VPN.pdf

Kaspersky Lab. (2005). *Worm.SymbOS.Cabir.a.* Retrieved July 12, 2005, from http://www.viruslist.com/en/viruslist.html?id=1689517

Table A9. Download speed measurements (download times for a 4.92 MB file)

With "Pointsec for Symbian OS"		**Without "Pointsec for Symbian OS"**	
WLAN (11 Mbit/s)	***GPRS (56 Kbit/s)***	***WLAN (11 Mbit/s)***	***GPRS (56 Kbit/s)***
1 min	16 min 33 s	59.62 s	21 min 1 s
57 s	20 min 54 s	58.28 s	19 min 52 s
59 s	18 min 16 s	59.14 s	18 min 2 s
1 min 2 s	19 min 25 s	1.0 min	18 min 42 s
1min 3 s	20 min 40 s	59.61 s	18 min 14 s
59.2 s	19 min 48 s	1 min 1 s	19 min 3 s
Average 60.03 s	**Average** 19 min 16 s	**Average** 59.61 s	**Average** 19 min 9 s
Standard Deviation 2.174093 s	Standard Deviation 97.912206 s	Standard Deviation 3.253739 s	Standard Deviation 67.337954 s
		Average +0.42 s with Pointsec	**Average +7 s** with Pointsec

Table A10. Connection time measurements (to mailbox on e-mail server)

With "Pointsec for Symbian OS"		**Without "Pointsec for Symbian OS"**	
WLAN (11 Mbit/s)	***GPRS (56 Kbit/s)***	***WLAN (11 Mbit/s)***	***GPRS (56 Kbit/s)***
31.62 s	36.96 s	24.29 s	35.91 s
32.23 s	40.70 s	25.16 s	32.07 s
32.17 s	37.11 s	26.21 s	33.88 s
31.86 s	38.42 s	25.34 s	31.30 s
31.19 s	39.71 s	25.45 s	43.76 s
32.42 s	42.51 s	25.46 s	30.14 s
Average 31.915 s	**Average** 39.235 s	**Average** 25.31833 s	**Average** 34.51 s
Standard deviation 0.454962 s	Standard deviation 2.165777 s	Standard deviation 0.618948 s	Standard deviation 4.965360 s
		Average +6.60 s with Pointsec	**Average +4.72 s** with Pointsec

Table A11. Connection time measurements (to the Web site www.nokia.fi)

With "Pointsec for Symbian OS"		**Without "Pointsec for Symbian OS"**	
WLAN (11 Mbit/s)	***GPRS (56 Kbit/s)***	***WLAN (11 Mbit/s)***	***GPRS (56 Kbit/s)***
22.68 s	58.68 s	21.55 s	27.17 s
26.70 s	57.65 s	22.27 s	28.98 s
27.62 s	59.06 s	23.24 s	29.20 s
31.47 s	56.06 s	24.58 s	29.76 s
27.87 s	59.03 s	22.94 s	30.43 s
25.14 s	58.23 s	23.31 s	30.11 s
Average 26.91333 s	**Average** 58.11833 s	**Average** 22.98167 s	**Average** 29.275 s
Standard deviation 2.942419 s	Standard deviation 1.140341 s	Standard deviation 1.028308 s	Standard deviation 1.165345 s
		Average + 4.93 s with Pointsec	**Average +28.84 s** with Pointsec

Table A12. Connection time measurements (to a SSH server)

With "Pointsec for Symbian OS"		**Without "Pointsec for Symbian OS"**	
WLAN (11 Mbit/s)	***GPRS (56 Kbit/s)***	***WLAN (11 Mbit/s)***	***GPRS (56 Kbit/s)***
<1 s	4.85 s	<1 s	4.9 s
<1 s	4.65 s	<1 s	4.71 s
<1 s	4.62 s	<1 s	4.13 s
<1 s	4.79 s	<1 s	4.61 s
<1 s	4.86 s	<1 s	4.42 s
<1 s	4.80 s	<1 s	4.52 s
Average <1 s	**Average** 4.761667 s	**Average** <1 s	**Average** 4.548333 s
Standard deviation < 1 s	Standard deviation 0.102258s	Standard deviation <1 s	Standard deviation 0.263015 s
			Average +0.21 s with Pointsec

Paraben Corporation. (2007). *Device seizure v1.2: Cell phone & PDA forensic software.* Retrieved September 27, 2007, from http://www.paraben-forensics.com/catalog/product_info.php?cPath=25&products_id=405

PC GuardianTechnologies. (2005). *PDASecure—Powerful security. Simple to use. Superior service & support.* Retrieved July 9, 2005, from http://www.pcguardiantechnologies.com/PDASecure/PDASecure_Brochure.pdf

Romsey Associates Ltd. (2005). *Details—PDALok V1.0..., PDALok—The technology behind the security software.* Retrieved July 3, 2005, from http://www.pdalok.com/pda_security_products/PDALok_details.htm

ROSISTEM. (2005). *Biometric education » Fingerprint.* Retrieved July 9, 2005, from http://www.barcode.ro/tutorials/biometrics/fingerprint.html

sfr Gesellschaft für Datenverarbeitung mbH. (2005). *The patented technology in all of our visual key products.* Retrieved July 3, 2005, from http://www.viskey.com/viskeypalm/index.html and http://www.viskey.com/tech.html

Sundaresan, H. (2003). *OMAP™ platform security features.* White Paper. Retrieved July 1, 2005, from http://focus.ti.com/pdfs/wtbu/omapplatformsecuritywp.pdf

Symantec Corporation. (2005). *Symantec AntiVirus™ for handhelds annual service edition, Symantec antiVirus for handhelds safeguards Palm and Pocket PC mobile users.* Retrieved July 9, 2005, from http://www.symantec.com/sav/handhelds/ and http://www.symantec.com/press/2003/n030825.html

Taylor, L. (2004). *Handheld security, part III: Evaluating security products.* Retrieved July 1, 2005, from http://www.firewallguide.com/pda.htm

The Familiar Project. (2005). *The familiar project.* Retrieved July 10, 2005, from http://familiar.handhelds.org/

Topaz Systems Inc. (2005). *Electronic signature pad with interactive LCD display.* Retrieved July 7, 2005, from http://www.topazsystems.com/products/specs/TL462.pdf

Trust Digital. (2005). *Trust digital 2005™.* Retrieved July 5, 2005, from http://www.trustdigital.com/downloads/productsheet_trust2005.pdf

Veridt, LLC. (2005). *Veridt verification and identification, product information.* Retrieved September 27, 2007, from http://www.veridt.com/Products.html

Villamil, F. (2005). *Firewall wizards: Looking for PDA firewall.* Retrieved July 10, 2005, from http://seclists.org/lists/firewall-wizards/2004/Jan/0031.html

This work was previously published in Mobile Multimedia Communications: Concepts, Applications, and Challenges, edited by G. Karmakar and L. Dooley, pp. 248-296, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.9
Multimedia Authoring for Communities of Teachers

Agnès Guerraz
INRIA Rhône-Alpes, France

Cécile Roisin
INRIA Rhône-Alpes, France

Jan Mikáč
INRIA Rhône-Alpes, France

Romain Deltour
INRIA Rhône-Alpes, France

ABSTRACT

One way of providing technological support for communities of teachers is to help participants to produce, structure and share information. As this information becomes more and more multimedia in nature, the challenge is to build multimedia authoring and publishing tools that meet requirements of the community. In this paper, we analyze these requirements and propose a multimedia authoring model and a generic platform on which specific community-oriented authoring tools can be realized. The main idea is to provide template-based authoring tools while keeping rich composition capabilities and smooth adaptability. It is based on a component-oriented approach integrating homogeneously logical, time and spatial structures. Templates are defined as constraints on these structures.

INTRODUCTION

We are involved in a multidisciplinary project, the aim of which is to support the activities of communities of practice (CoP) in pedagogical environment. This project will provide tools for document production and for document reuse in heterogeneous applications. The objective is to reduce the current limitations caused by the proliferation of data sources deploying a variety

of modalities, information models, and encoding syntaxes. This will enhance applicability and performances of document technologies within pedagogically consistent scenarios. In this paper, we will focus on the authoring needs of teacher communities and propose a new authoring model, LimSee3.

In the educational context, there exists a large variety of authoring tools, see (Brusilowski, 2003) for an extensive review. The main objective of these systems is to provide adaptive educational hypermedia thanks to well-structured hyperlinked content elements that are mostly static content. In Hoffman and Herczeg (2006), the created documents are video centric, providing a way to add timed hot-spot embedding additional media and interaction facilities in the resulting hypervideo. The time structure is, therefore, straightforwardly given by the video media, while the time model of our approach (given by the SMIL time model) is much more general. In our project, we want to provide educators with a way to take advantage of multimedia synchronization to offer more lively pedagogical material. But it is worth noting that multimedia brings a higher order of complexity for authors. In order to reduce this complexity, we propose a multimedia authoring model that will provide similar authoring services than formed-based hypermedia systems (Grigoriadou & Papanikolaou, 2006).

The LimSee3 project aims at defining a document model dedicated to adaptive and evolutive multimedia authoring tools, for different categories of authors and applications, to easily generate documents in standard formats. Our approach is to focus on the logical structure of the document while keeping some semantics of proven technologies such as SMIL (SMIL). This provides better modularity, facilitates the definition of document templates, and improves manipulation and reusability of content. The LimSee3 authoring process is given on Figure 1: a document is created from a template by adding content in an application-guided way. The obtained LimSee3 document can be exported into one or several presentation documents suitable for rendering.

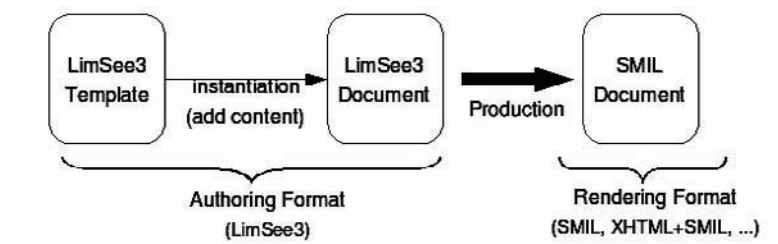
This paper is organized as follows: next section presents a scenario example that will be developed throughout the paper and thereby analyzes CoPs requirements for authoring multimedia documents. We then define the main concepts on which multimedia authoring tools are based, and we classify existing approaches in the light of these concepts. After that, we introduce the LimSee3 document model and show how it can be used for the development of authoring tools tuned for specific CoPs. The last section presents the current state of our development and our perspectives.

A LEARNING-ORIENTED EXAMPLE OF AUTHORING

Multimedia Storytelling for Enhanced Learning

Educators have integrated practice into their curriculum to different degrees; Figure 2 shows this continuum and how LimSee3 can be

Figure 1. The authoring process in LimSee3



naturally used to enhance authoring multimedia documents.

Edward Bilodeau (2003) illustrated that moving towards full immersion requires substantial changes to course design. Careful consideration must be given to the optimal location for student learning to occur on this continuum. Using templates in LimSee3 authoring tool for pedagogical approach allows production process during this continuum. It gives a way of making things simpler and faster to teachers and writers. It focuses on pedagogical issues. It produces practical units of learning (UoL).

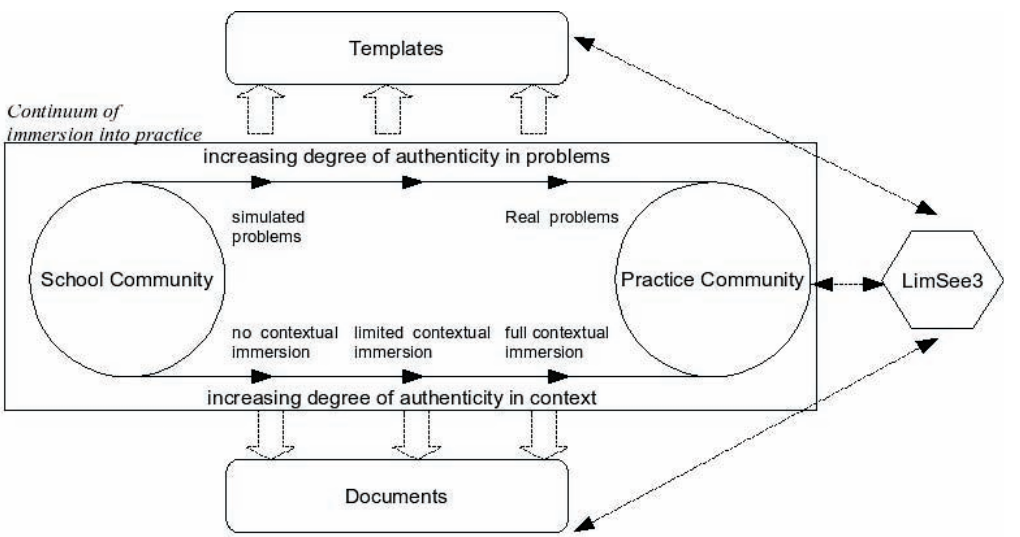
Researchers such as Dolores Durkin (1961), Margaret Clark (1976), Regie Routman (1988; 1991), and Kathy Short (1995) have found evidence that children who are immersed in rich, authentic literary experiences become highly engaged in literature and develop literary awareness. Their studies revealed that positive and meaningful experiences with books and written language play a critical role in the development of literacy skills. Other researchers have found that students acquired reading and thinking strategies in literature-based programs that included teacher-led comprehension instruction (Baumann, 1997; Block, 1993; Goldenberg, 1992/1993).

Storytelling in the Learning Process

Stories are basic: we start telling our children about the world by telling them stories. Stories are memorable: their narrative structure and sequence makes them easy to remember. "What happens next?" is a very basic form of interest and engagement. Stories are everywhere: very rarely in real life do we set out to convey ideas in terms of hierarchies or taxonomies or bullet points. Instead, we tell stories. Teaching is one of the predominantly professional activities that do habitually communicate by means of stories, and also use elaborated language codes.

We want to go deeper inside the multimedia process, taking into account the advantage of creating multimedia in the immersion process. As an example, U-Create (Sauer, Osswald, Wielemans, & Stifter, 2006) is an authoring tool centered on 2-D and 3-D scenes targeted to nonprogrammers who want to easily produce story documents. The tool is based on predefined structural elements

Figure 2. Continuum of immersion into practice (Adopted from Hogan, 2002) and LimSee3 use



(from story, scene, to action, stageSet, and asset) and associated dedicated GUIs.

In this paper we consider a group of teachers that are working together—a CoP in our terminology—to create and share course materials based on tale storytelling.

TALE LEARNING EXAMPLE

Little Red Riding Hood Example

Little Red Riding Hood is a well-known and well-loved fairytale dating back to over 3 centuries ago when it was first published by Charles Perrault in his Histoires ou Contes du temps passé in 1697 and based on oral European tales. Since then, Little Red Riding Hood has been retold in a variety of forms and styles, as Big Books and Lift-the-flap books, as poems and plays, and whilst some details may have changed, many of the essential elements have stayed the same. Little Red Riding Hood makes a great literature teaching unit theme for elementary school.

A general synopsis follows:

Act I.1

Setting at the edge of a forest; the Little Red Riding Hood goes off to take a basket to her ill grandmother, her mother warns her not to dawdle in the woods or to talk to strangers.

Act I.2

A place inside the forest. Woodcutters can be heard chopping wood. Little Red Riding Hood comes out of some bushes. As she pauses to pick some flowers, the wolf catches sight of her. On the path he stops her and makes up a story about a shortcut to grandmother's house. When he challenges her to see who will get there first, she agrees, and both of them run off in different directions as the woodcutters resume their work.

Act II.1

The chorus explains that the wolf has not eaten for 3 days and was able to get to grandmother's house first. The wolf, pretending to be Little Red Riding Hood, manages to get into the house and swallow grandmother. He takes her place in the bed before Little Red Riding Hood arrives. In several questions she expresses her surprise at how differently grandmother looks now, and the wolf swallows her.

Act II.2

Some hunters and woodcutters, who have been tracking the wolf, come by and enter the house. They find the wolf asleep and open his belly to let grandmother and Little Red Riding Hood out. After they sew up the wolf again, he repents and is permitted to live in the forest as long as he lives up to his promise to be good.

In the learning process, it is possible to exploit this story in different approaches (Franc-parler. org, 2006), for instance:

Story and the time: a) After reading the tale and giving explanations of difficult points and incomprehensions, to work on the chronology from drawings. b) Try and feel the knowledge of the terms "before," "later," during this time line.

Oral expression: a) Drawing images and explaining them. b) Playing dialogues without written support. c) Reciting a rhyme or poem, singing a song.

A variety of resources, ranging from texts, illustrations, media presentations to computer-based interactive materials for students are available for use in classroom. Based on these materials, a teacher can propose:

- **Story:** Tell the story, watch and comment movies
- **Songs:** Organize some spoken drill type activities
- **Handcraft activities:** Propose drawing, folding, coloring of sceneries, puppets, and so forth

- **Play:** Study and put on stage a personalized version of the story

Basically, the units of learning that are exchanged in this CoP of teachers are multimedia story documents that are composed of sequences of story steps where data elements are heterogeneous and multimedia. The challenges are to enrich information with the synchronization of data elements (for instance an activity with the corresponding material) and to provide a document structure enabling knowledge sharing and reusability (of stories).

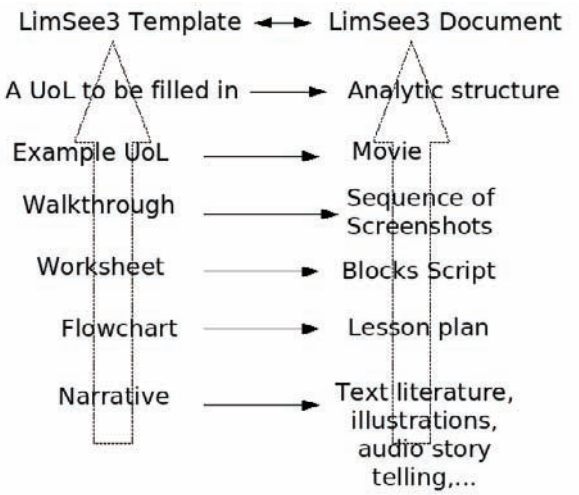
The CoP of teachers needs templates for making things simpler, faster, for being focused on pedagogical issues, for producing practical units of learning. The Fig. 3 shows the structural link between LimSee3 Template and LimSee3 document contents. At the lower level, a narrative part inside the template corresponds to text literature and/or illustration and/or audio storytelling. At a higher level, a template walkthrough corresponds to a sequence of screenshots. The first level is offered by the BNF (BNF, 2001) that, for instance, gives out textual contents and illustrations. To fully instantiate upper levels, we show a possible making of the tale with a logic modeling [template and document] from which we can extract levels to enhance associated authoring.

From the continuum of immersion into practice, represented inside the rectangle of the Fig. 2, we learn that the greater the degree of authenticity of the learning activities, the more the students will be able to be integrated into the practice. Different programs and courses benefit from different levels of immersion. Moving towards full immersion requires substantial changes to course design. Teachers need authoring tools to set up these types of pedagogical materials. Careful consideration must be given to the optimal location for student learning to occur in this continuum. In this CoP, a number of teachers will create templates to build this type of very specialized tools. Such a modeling will naturally emerge from CoP work, using LimSee3 inside this continuum (see Fig. 2).

Basic Requirements for CoP-Oriented Learning

In order to be useful, the cooperative services to be provided to the CoPs must have the two following basic features: (i) authoring tool of stories dedicated to teachers; (ii) access tool

Figure 3. LimSee3 template and document links



to read the existing stories. Looking more closely at the ways in which CoPs participants are producing multimedia information, we can identify some requirements for the authoring and presentation platform:

1. Simple and efficient authoring paradigms—because CoPs members are not (always) computer science technicians.
2. Easy and rapid handling of the authoring tool—because new members can join CoPs.
3. Modular and reusable content—because multimedia information results in a coconstruction process between members.
4. Evolutive structuring of documents —because of the dynamic nature of CoPs objectives.
5. Use of standard formats—because CoPs need portability, easy publishing process, and platform independence.

Basically, our approach proposes a template mechanism to cope with requirements 1 and 2, a component-based structuring enabling requirements 3 and 4, and relies on proven standard technologies to ensure the last requirement. Before further stating our authoring model, we present in the next section the main concepts and approaches of multimedia authoring on which this work is based.

MULTIMEDIA DOCUMENTS AND MULTIMEDIA AUTHORING

In traditional text-oriented document systems, the communication mode is characterized by the spatial nature of information layout and the eye's ability to actively browse parts of the display. The reader is active while the rendering itself is passive. This active-passive role is reversed in audio-video communications: active information flows to a passive listener or viewer. As multimedia documents combine time, space, and interactivity, the reader is both active and passive. Such documents contain different types of elements such as video, audio, still-picture, text, synthesized image, and so on, some of which have intrinsic duration. Time schedule is also defined by a time structure synchronizing these media elements. Interactivity is provided through hypermedia links that can be used to navigate inside the same document and/or between different documents.

Due to this time dimension, building an authoring tool is a challenging task because the WYSIWYG paradigm, used for classical documents, is not relevant anymore: it is not possible to specify a dynamic behavior and to immediately see its result. Within the past years, numerous researchers have presented various ways of authoring multimedia scenarios, focusing on the understanding and the expressive power of synchronization between media components: approaches can be classified in absolute-based (Adobe, 2004), constraint-based (Buchanan & Zellweger, 1993; Jourdan et al., 1998), event-based (Sénac, Diaz, Léger, & de Saqui-Sannes, 1996) and hierarchical models (SMIL), (Van Rossum, Jansen, Mullender, & Bulterman, 1993). Besides, to cope with the inherent complexity of this kind of authoring, several tools (Adobe, 2004), (Microsoft, n.d.), (Hua, Wang, & Li, 2005) have proposed limited but quite simple solutions for the same objective. Dedicated authoring, template-based authoring, and reduced synchronization features are the main techniques to provide reasonable editing facilities. But we can notice that these tools generally also provide scripting facilities to enrich the authoring capabilities and therefore, lose in some way their easiness.

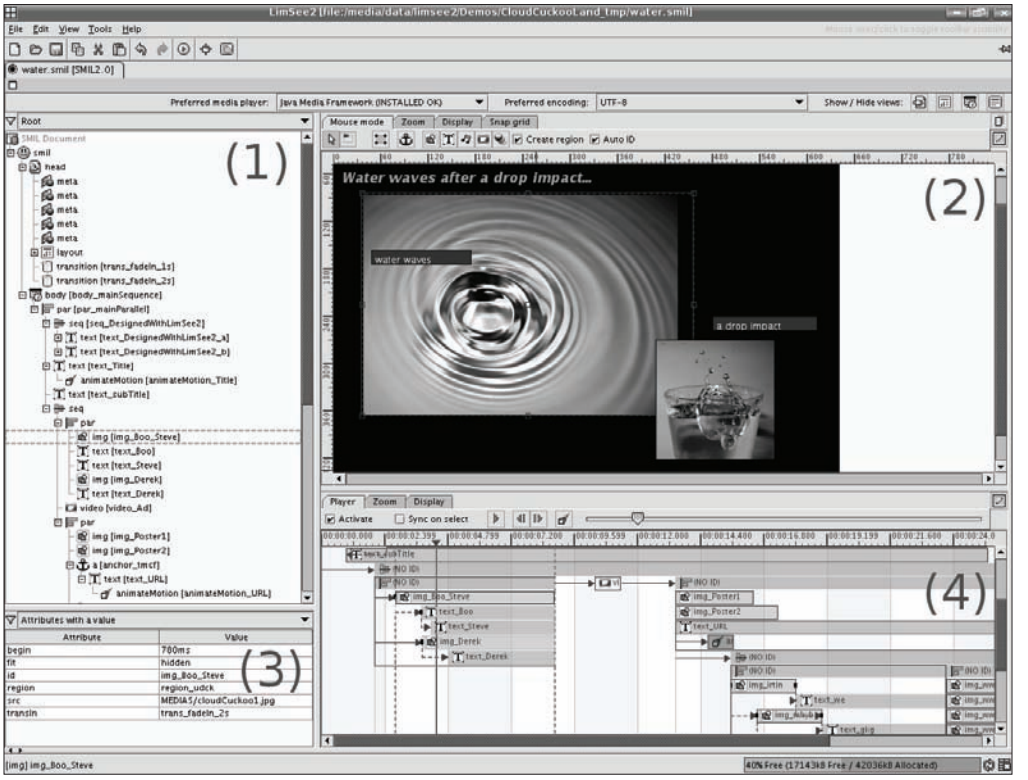
Beside timelines, script languages, and templates, intermediate approaches have been proposed through "direct manipulation" and multiviews interface paradigms. IBM XMT authoring tool (IBM) and SMIL tools such as LimSee2 (LimSee2) and Grins (Oratrix GRiNS) are good examples. In LimSee2, the time structure of SMIL is represented for instance in a hierarchical

timeline as shown in of Fig. 2 (4). Time bars can be moved or resized to finely author the timing scenario. This kind of manipulation has proven very useful to manipulate efficiently the complex structures representing time in multimedia XML documents.

However, even if XMT and SMIL are well-established languages, these tools are too complex for most users because they require a deep understanding of the semantics of the language (e.g., the SMIL timing model). Moreover, these models generally put the time structure at the heart of the document, whereas it does not always reflect exactly the logical structure in the way it is considered by the author. Our approach instead sets this logical dimension as the master structure of the document, which is a tree of modular components, each one specifying its own time and spatial structures. Additionally, the document can be constrained by a dedicated template mechanism.

A template document is a kind of reusable document skeleton that provides a starting point to create document instances. Domain-specific template systems are a user-friendly authoring solution but require hardly extensible dedicated transformation process to output the rendering format. We chose, on the contrary, to tightly integrate the template syntax in the document: the template is itself a document constrained by schema-like syntax. The continuum between both template and document permits to edit templates as any other document, within the same environment, and enables an evolutive authoring of document instances under the control of templates. There is no need to define a dedicated language to adapt to each different use case.

Figure 4. Multiview authoring in LimSee2



We believe that the combination of document structuring and template definition will considerably help CoPs in (i) reusability of materials, (ii) optimization of the composition and life cycle of documents, (iii) development and transmission of knowledge, (iv) drawing global communities together effectively.

THE LIMSEE3 AUTHORING LANGUAGE

Main Features

In the LimSee3 project, we define a structured authoring language independently of any publication language. Elements of the master structure are components that represent semantically significant objects. For instance, a folktale can be seen as a sequence of scenes. Each step is composed of several media objects and describes a phase of the story (departure from home, encountering the wolf,...). Components can be authored independently, integrated in the document structure, extracted for reusability, constrained by templates, or referenced by other components.

The different components of a multimedia document are often tightly related with one another: when they are synchronized or aligned in space, when one contains an interactive link to another, and so on. Our approach, which is close to the one proposed in Silva, Rodrigues, Soares, and Muchaluat Saade (2004), is for each component to abstract its dependencies to external components by giving them symbolic names. This abstraction layer facilitates the extraction of a component from its context; thus, enhancing modularity and reusability.

Finally, the goal is to rely on existing proven technologies, in both contexts of authoring environments and multimedia representation. The timing and positioning models are wholly taken from SMIL. Using XML (XML, 2006) provides excellent structuring properties and enables the use of many related technologies. Among them are XPath (Xpath, 1999), used to provide fine-grained access to components, and XSLT (XSLT, 1999), used in templates for structural transformation and content generation.

The authoring language is twofold: it consists in a generic document model for the representation of multimedia documents, and it defines a dedicated syntax to represent templates for these documents.

In this section, we describe the main features of the LimSee3 language and we illustrate their syntax with short excerpts of the storytelling example.

Document Model

A *document* is no more than a document element wrapping the root of the object hierarchy and a head element containing metadata. This greatly facilitates the insertion of the content of a document in a tree of objects, or the extraction of a document from a subtree of objects.

A compound object is a tree structure composed of nested objects. Each compound object is defined by the object element with the type attribute set to compound. It contains a children element that lists child objects, a timing element that describes its timing scenario, a layout element that describes its spatial layout, and a related element that abstracts out dependencies to external objects.

The value of the localId attribute uniquely identifies the component in the scope of its parent object; thereby also implicitly defining a global identifier id when associated with the localId of the ancestors. In Example 1, the first child of object scene1 has the local id title and hence is globally identified as scene1.title.

The timing model, and similarly the positioning model, is taken from SMIL 2.1. The timing element defines an SMIL time container. The timing scenario of a component is obtained by composition of the timed inclusions defined by

the timeRef elements, whose refId attributes are set to local ids of children.

A *media object* is actually a simple object that wraps a media asset, i.e. an external resource (such as an image, a video, an audio track, a text...) referenced by its URI. It is defined by the object element with the type attribute set to either text, image, audio, video, or animation. The URI of the wrapped media asset is put into the src attribute. Example 2 shows an image media

Figure 5. Example 1 - A simple scene LimSee3 document

```
<document xmlns="http://limsee3.gforge.inria.fr/ns"
  xmlns:template=".../ns/template"
  xmlns:smil="http://www.w3.org/2005/SMIL21/">
  <head><!-- some metadata --></head>
  <object localId="scene1" type="compound">
    <children>
      <object type="text" localId="title">
        ...
      </object>
      <object type="image" localId="illustration1">
        ...
      </object>
      ...
      <object type="compound"
        localId="navigation-bar">
        ...
      </object>
    </children>
    <timing>
      <smil:par dur="30min">
        <timingRef refId="title" />
        <timingRef refId="illustration1" />
        ...
        <timingRef refId="navigation-bar" />
      </smil:par>
    </timing>
    <layout height="600" width="800" />
  </object>
</document>
```

Figure 6. Example 2 - A LimSee3 object with an external dependency relation

```
<object localId="right-button" type="image"
  src="./medias/right-arrow.png">
  <related>
    <ref localId="target" refId="story.scene2"/>
  </related>
  <children>
    <object type="area" localId="link">
      <attribute name="src">#<value-of
        refName="target" select="@id" />
    </object>
  </children>
  ...
</object>
```

object with local id right-button that wraps the media asset identified by the relative URI ./medias/right-arrow.png.

Area objects inspired from the SMIL area element can be associated to media objects. They are used for instance to structure the content of a media object or to add a timed link to a media object. An area is defined as an object element with the type attribute set to area. For instance, in Example 2 the media object right-button has a child area, which defines a hyperlink.

Relations of dependency between objects are described independently of their semantics in the document. External dependencies are declared with ref elements grouped inside the related child element of objects. The value of refId of a ref element is the id of the related element, and the value of localId is a symbolic name that is used within the object to refer to the related object. For instance, in Example 2, object right-button provides a clickable image that links to the object story.scene2 by first declaring the relation in a ref element and then using this external object locally named target to set the value of the src attribute of the link, using attribute and value-of elements taken form XSLT.

Templates

Template nodes aim at guiding and constraining the edition of the document. In order to have better control and easy GUI setup, the language includes two kinds of template nodes: media placeholders and repeatable structures.

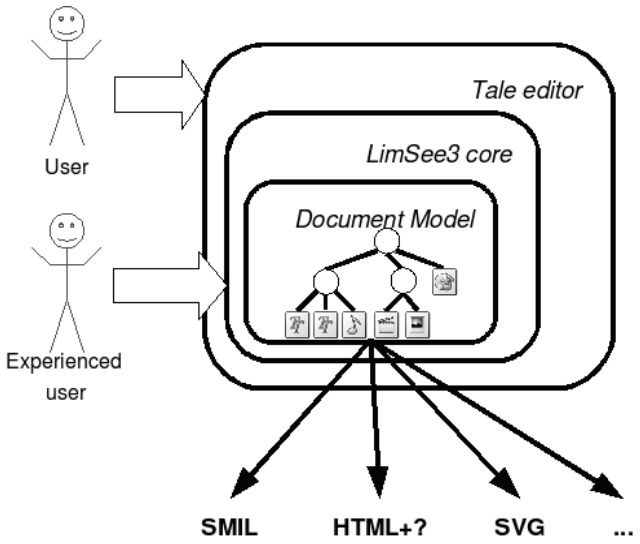
A *placeholder* is a template node that defines a reserved place for a media object. It is represented by an object element, whose type and src attributes are not (yet) set. It specifies the kind of media resources it accepts in a special template:types attribute (the values can be text, img, audio, video, animation, or a list of these types). The author can also specify content that will be displayed to invite the user to edit the media zone with the template:invite element (of any media type). For instance, Example 3 shows a media placeholder title for a text, with textual invitation. During the authoring process, placeholders are filled with media objects inserted by the user.

A *repeatable structure*, represented by the template:model element, is a possibly complex template node that defines a homogeneous list of objects. Each item of the list matches the model. The cardinality of the list can be specified with the min and max attributes. Example 3 shows a tale

Figure 7. Example 3 - A scene template

```
<template:model name="tale-scene" min="1">
  <object>
    <children>
      <object localId="title" template:types="text">
        <children>
          <template:invite type="text">Fill in the title of this scene</template:invite>
        </children>
      </object>
      <template:model name="illustrations" min="1" max="5">
        ...
      </template:model>
      <object localId="questions" template:types="text">...</object>
      <object localId="navigation" type="complex">...</object>
    </children>
    <timing>...</timing>
    <layout>...</layout>
  </object>
</template:model>
```

Figure 8. The LimSee3 three-layer architecture



scene template named tale-scene: this complex model is composed of several placeholders (title, questions), an embedded model (illustrations), and the navigation object.

Finally, our model makes it is possible to lock out parts of a document with the locked attribute, to prevent authors from editing these parts. This allows for instance to guide more strongly inexperienced users by restricting their access to the only parts of the document that make sense to them.

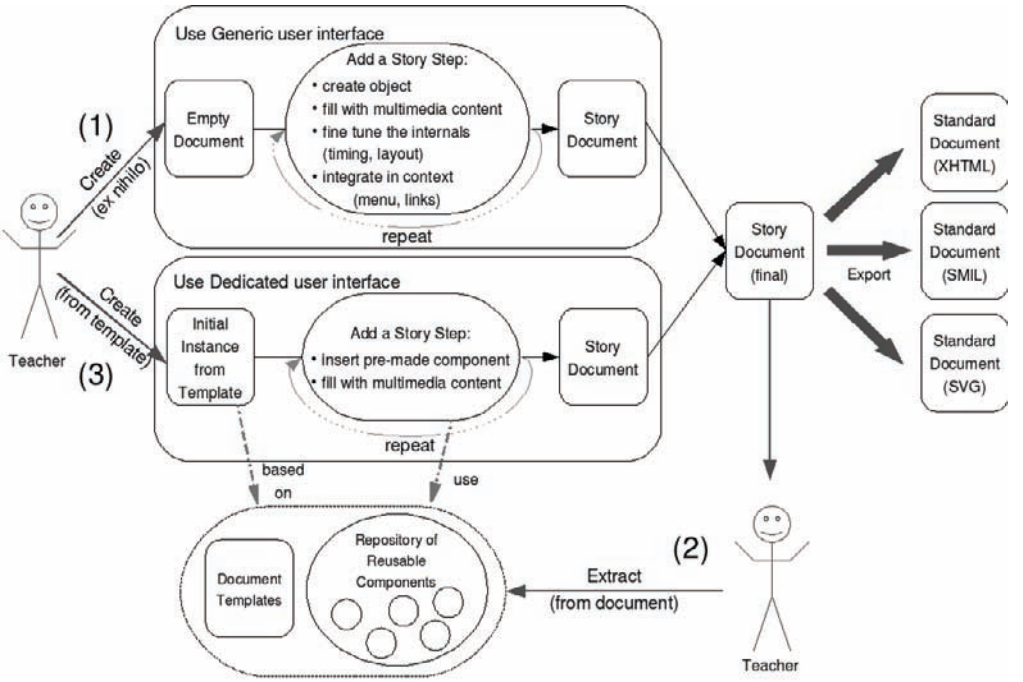
AUTHORING WITH LIMSEE3

The LimSee3 model defined in the previous section aims at being an internal kernel of our authoring system as can be seen on Figure 8. This model is hidden from the user by several abstraction layers: experienced users interact with a full-featured generic platform (LimSee3 core) that enables them to finely tune all document properties, however, at the cost of some technical overhead, while basic users interact with template-specific simplified interfaces that allow them to author documents with less effort. In the case of a teacher's CoP that wants to create and share multimedia materials for tale telling, the benefits that could be achieved with LimSee3 are the following:

- **To describe the author's vocabulary** by structuring basic medias into author-defined logical multimedia structures ("tale scenes" viewed as collections of "illustrations" and "narrations," rather than mere "pictures" and "texts")
- **To adopt the author's vocabulary in the authoring process** by leveraging the logical coherence between the document to produce and the way to produce it (the template structure reflects the logics of the presentation, not its technical needs)
- **To facilitate document reuse in the CoP** by easily extracting, adapting, merging documents, applying alternate layouts for different pedagogical purposes

Figure 9 shows the different steps of the production of tale tellings by the member of teachers CoP. First, an experienced author creates a tale story from scratch using the LimSee3 core plat-

Figure 9. Authoring with LimSee3



form (flow (1) in the Figure 9) in order to define the logical structure of this pedagogical material that will allow a fruitful use in classroom. Eventually this multimedia tale will be refined thanks to inputs from other teachers of the CoP. When a consensus is reached, this teacher can use the LimSee3 core to extract a template document from this instance (flow (2) in the Figure 9). The main structure of the document, in this case a sequence of scenes, can be constrained by template nodes such as repeatable structures. The result of this step will be a dedicated authoring interface that other teachers can use (flow (3) in the Figure 9) to create new multimedia tale stories. This is a typical example of participative design leading to the development of a dedicated tool based on the LimSee3 generic platform.

The Figures 10 and 11 illustrate this last step of authoring with a dedicated GUI:

- Figure 10 shows how the placeholders defined in the template structure can lead to simple drag-and-drop authoring actions.
- Figure 11 illustrates the advantages gained with the separation of logical, spatial, and time information. It allows the authoring and rendition of several scenarios of the same content: thanks to a direct manipulation in the timeline view, an author has defined a sequential display of the illustrations instead of the default parallel one.

Finally, the proposed application can evolve to take into account new needs of the CoP members. For instance, a teacher wants to register his/her course, using a camera that films her/him while (s)he gives a talk illustrated with the multimedia tale document. In order to easily synchronize the video with the different parts of the tale document, the authoring tool is enriched with a simple control panel, as can be seen on the left part of Figure 12.

Figure 10. Instantiating a template document by drag-and-drop

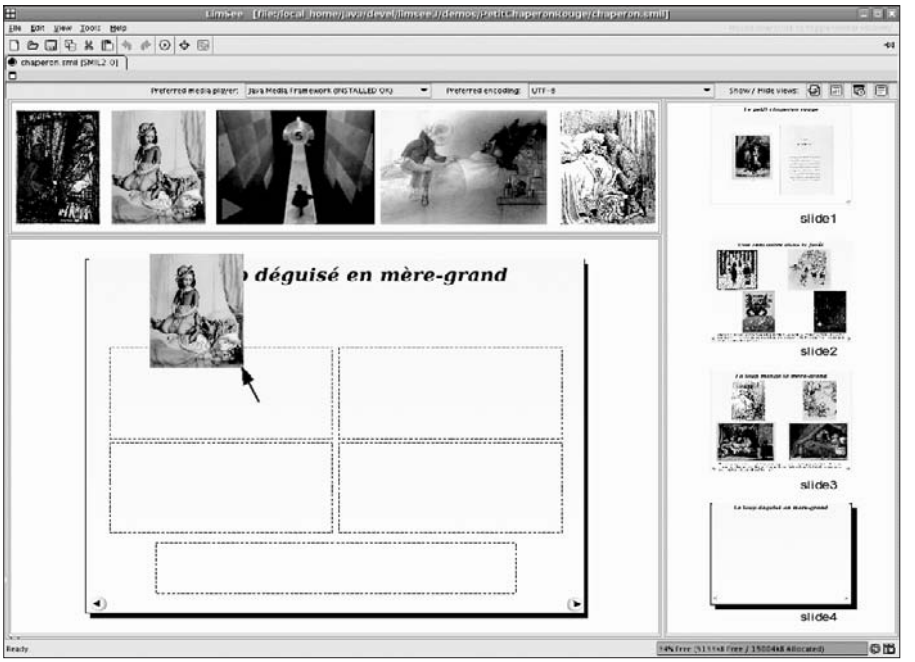


Figure 11. Modifying the timeline

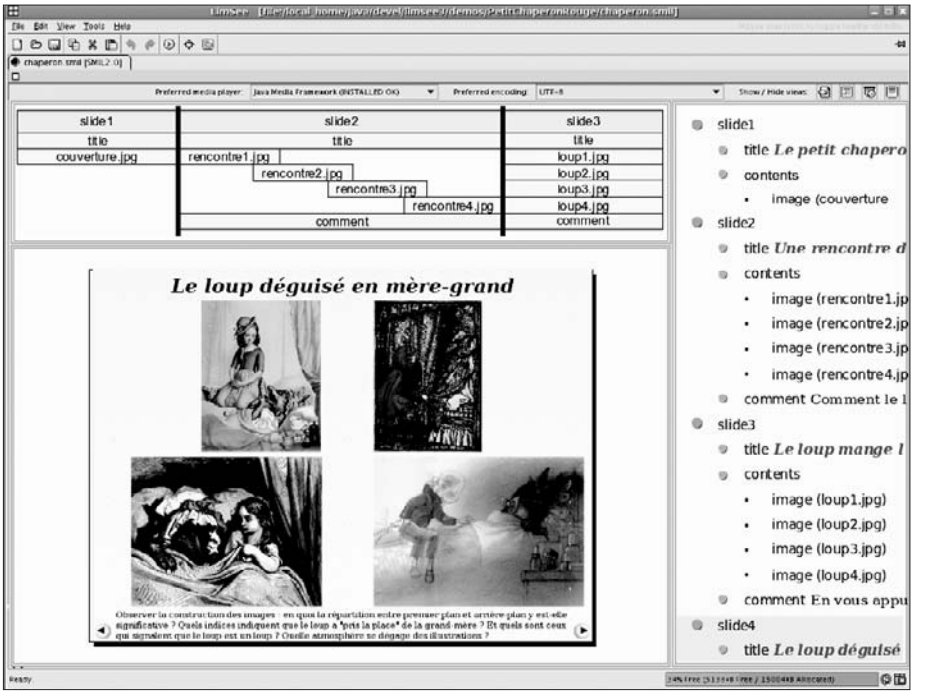


Figure 12. Synchronizing a video with a story document

CONCLUSION

The LimSee3 model leads to the development of authoring tools that fit the requirements stated at the beginning of the paper. The LimSee3 core is currently under development as a cross-platform open-source Java software: we provide this generic platform with widgets to manipulate all the elements defined in the model (documents, compound objects, timing and layout details, relations...). It provides features based on the proven authoring paradigms such as multiviews, timeline, structure tree, and 2-D canvas.

The model presented in this paper develops a practice-based approach to multimedia authoring dedicated to communities where collaborative and participative design is of high importance. It improves reusability with template definitions and with the homogeneous structuring of documents. In the context of Palette, we will use this model to develop dedicated authoring tools for pedagogical CoPs.

REFERENCES

Adobe. (2004). Adobe Authorware 7 and Macromedia Director MX 2004. Retrieved from http://www.adobe.com/products/

Baumann, J. F. & Ivey, G. (1997). Delicate Balances: Striving for curricular and instructional equilibrium in a second-grade, literature/ strategy-based classroom. *Reading Research Quarterly, 32*(3), 244.

Bilodeau, E. (2003, November 7-8, 2003). *Using communities of practice to enhance student learning.* Paper presented at EGSS Conference 2003, McGill University, Montreal.

Block, C. C. (1993). Str-ategy instruction in a literature-based reading program. *Elementary School Journal, 94*(2), 139-151.

BNF. (2001). *Autour du Petit Chaperon rouge.* Retrieved from http://expositions.bnf.fr/contes/pedago/chaperon

Brusilovsky, P. (2003). Developing adaptive educational hypermedia systems: From design models to authoring tools. In *Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive and intelligent educational software*. Kluwer Academic Pub.

Buchanan, M. C., & Zellweger, P. T. (1993). *Automatic temporal layout mechanisms* (pp. 341-350). ACM Multimedia.

Clark, M. (1976). *Young fluent readers: What can they teach us?* Heinemann.

Durkin, D. (1961). Children who read before grade one. *The Reading Teacher. 14*, 163-166.

Franc-parler.org. (2006). La communauté mondiale des professeurs de français. Franc-parler.org un site de l'Organisation internationale de la Francophonie. Retrieved from http://www.francparler.org/parcours/conte.htm

Goldenberg, C. (1992/1993). Instructional conversations: promoting comprehension through discussion. *The Reading Teacher, 46*(4), 316-326.

Grigoriadou, M., & Papanikolaou, K. (2006). Authoring personalised interactive content. In Proceedings of *First International Workshop on Semantic Media Adaptation and Personalization (SMAP'06)*, (pp. 80-85).

Hoffmann, P., & Herczeg, M. (2006). *Interactive narrative systems - Hypervideo vs. storytelling integrating narrative intelligence into hypervideo* (LNCS 4326, pp. 37-48).

Hogan, K. (2002). Pitfalls of community-based learning: How power dynamics limit adolescents' trajectories of growth and participation. *Teachers College Record, 104*(3), 586-624.

Hua, X., Wang, Z., & Li, S. (2005). *LazyCut: Content-aware template based video authoring.* ACM Multimedia.

IBM. (n.d.). Authoring in XMT. Retrieved from http://www.research.ibm.com/mpeg4/Projects/AuthoringXMT/

Jourdan, M., Layaïda, N., Roisin, C., Sabry-Ismail, L., & Tardif, L., Madeus (1998, September). An authoring environment for interactive multimedia documents. In ACM Multimedia, Bristol, UK, September 1998, pp. 267272.

Jourdan, M., et al. (1998). *Madeus, an authoring environment for interactive multimedia documents.* ACM Multimedia.

LimSee2. (2003-2006). Retrieved from http://limsee2.gforge.inria.fr/

Microsoft. (n.d.). *MS Producer for PowerPoint.* Retrieved from http://www.microsoft.com/office/powerpoint/producer/prodinfo/

Oratrix GRiNS. Retrieved from http://www.oratrix.com/

Palette. Retrieved from http://palette.ercim.org/

Routman, R. (1988). *Transitions: from literature to literacy.* Heinemann.

Routman, R. (1991). *Invitations: Changing as Teachers and Learners K-12*. Heinemann.

Sauer, S., Osswald, K., Wielemans, X., & Stifter, M. (2006). S*tory authoring - U-Create: Creative authoring tools for edutainment applications* (LNCS 4326, pp. 163-168).

Sénac, P., Diaz, M., Léger, A., & de Saqui-Sannes, P. (1996). Modeling logical and temporal synchronization in hypermedia systems. *IEEE Journal of Selected Areas on Communications, 14*(1), 84-103.

Short, K. (1995). *Research and professional resources in children's literature: Piecing a patchwork quilt.* International Reading Association.

Silva, H., Rodrigues, R. F., Soares, L. F. G., & Muchaluat Saade, D. C. (2004). *NCL 2.0:Iintegrating new concepts to XML modular languages.* ACM DocEng.

SMIL. (n.d.). *SMIL 2.1.* Retrieved from http://www.w3.org/TR/SMIL2/

Van Rossum, G., Jansen, J., Mullender, K., & Bulterman, D. C. A. (1993). *CMIFed:A presentation environment for portable hypermedia documents.* ACM Multimedia.

XML. (2006). *Extensible markup language (XML) 1.1.* Retrieved from http://www.w3.org/TR/xml11

XPath. (1999). *XML path language (XPath) 1.0.* Retrieved from http://www.w3.org/TR/xpath

XSLT. (1999). *XSL transformations 1.0.* Retrieved from http://www.w3.org/TR/xslt

This work was previously published in International Journal of Web-based Learning and Teaching Technologies, edited by L. Esnault, pp. 1-18, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.10
Screenspace

Kevin H. Jones
University of Oregon, USA

ABSTRACT

From tiny interactive cellphone screens (keitai) to supersized jumbo LED displays, Tokyo's urban landscape is changing drastically. A corner that once displayed billboards that occasionally flipped has now become lit-up and is in constant motion. Keitai, with their built-in cameras, now allow images to be sent from one to another and have become essential to urban life. As these screens become architectural and fashion statements, Tokyo's nomadic high-tech culture is commuting even greater distances, living in more compact housing, and allowing for "cellspace" and "screenspace" to merge.

Today we live in the imaginary world of the screen, of the interface…and networks. All our machines are screens, we too have become screens, and the interactivity of men has become the interactivity of screens.

–Jean Baudrillard

INTRODUCTION

Since the mid-1970s, with the development of the personal computer, interactive screens have slowly become part of our daily lives. The first interactive screens introduced were monochromatic, very much text-driven, and still referred to the structure of the printed page. Rudimentary images had to be cleverly created with the characters presented on the computer's keyboard.

Today, we find ourselves surrounded by vast wireless networks that we interface with through the use of electronic devices ranging from simple text messaging to high-resolution graphics on laptop computers. Various sized screens with rich GUI's (graphical user interface) are portals that allow us to interact with software on these devices. This invisible electromagnetic, hertzian space[1] has been growing rapidly since the mid-20th century. Urban landscapes are not only changing because of the mobility and interactivity of these screens, but also because of the proliferation of large-screen LED displays that are at every corner vying for our attention.

From Times Square in New York City to Shibuya Crossing in Tokyo, urban space is changing drastically. A corner that once displayed billboards that occasionally flipped has now become lit up and is in constant motion. Even though these screens are neither mobile nor interactive, the graphics that are displayed upon their surfaces light up the streets and mirror the motion of the city.

With this chapter, I will be looking at the proliferation of screens found in Tokyo and how they have transformed public and personal space. I will look closely at the "keitai"[2] with its tiny screens that dot Tokyo's landscape, all the way up to the "supersized" LED displays found in major entertainment and shopping areas in Tokyo. I will examine various screens' mobility, interactivity, social effects, and how graphics and typography are displayed on these devices.

CELLSPACE

Tokyo is a city in flux; it is in constant motion. With the daily influx of millions of commuters all with "keitai", cellspace is a dominant structure upon its urban fabric. The term "cellspace" was first coined by David S. Bennahum while using his Palm Pilot with a wireless modem to access the web and send e-mail while waiting for a NYC subway train in 1998. Cellspace surrounds us with its tangible and virtual interfaces, as well as the invisible wireless network that it has created. Today, it is everywhere and is a communication space for the nomad, allowing unparalleled movement and convenience.

Japan is a country with limited space, and Tokyo is one of the most crowded and expensive cities in the world. With high rents and one -third of living spaces averaging 400 square feet, urban apartments average approximately one room in size. Space is at a premium. Because of this lack of affordable and large housing in Tokyo, many Japanese commute up to four hours round trip from their homes in the surrounding suburbs to their places of employment. This distance from their offices makes for a very long commute on Tokyo's extensive rail system—perfect for the use of the popular keitai.

During my first residence in Japan in 1998, my daily one-way commute was one hour and 30 minutes from my front door to the Musashanita train station, then a train to the Megro station where I changed to the JR Yamanote line, which brought me to the Shinjuku train station, where again, I switched to another train that took me to the Hatsudai train station. After a 10-minute walk from there, I would finally arrive at the front door of the graphic design studio where I worked. I quickly discovered that this commute is not unusual for the Tokyo worker, as I was commuting with millions of others five days a week. During my daily commute, I would often read, just as all of the other passengers did to pass the time away and to avoid eye contact. But this has now changed: the majority of passengers are not reading the static text of a book or newspaper, but clicking away with their thumbs at their keitai keys. These commuters are not talking, (that would be impolite), but writing e-mails and entering into their individual cellspace.

The keitai that the Japanese use are extremely sophisticated: all current models are able to send and receive e-mail and have access to a limited amount of official webpages (with a subscription to one of the most popular service providers: I-mode or J-sky). Along with access to an already familiar interface, many phones allow low to medium resolution images to be captured with the keitai built-in camera and sent to another keitai or computer. Called sha-mail[3], these camera-enabled keitai have become extremely popular. Their screens allow thousands of colors of resolution and have built in 3D polygon engines. Streaming video from one keitai to another is on the horizon. With the rapid development of technologies related to the keitai, one can only imagine what the next generation will bring. How wonderful would it

be to conduct a videoconference with my associates at Sony as I finish my cup of coffee at the local coffee shop. Or, even better, play network games, watch bootlegged movies, or use it as a spy-camera. What this next generation of keitai will present is not only a hyper-cellspace: it will add to what I will refer to as screenspace.

SCREENSPACE

Tokyo is a city of not only cellspace, but of screenspace. Electronic displays are everywhere, from the tiny keitai screens to the monolithic LED screens found on the buildings in Shibuya, Shinjuku, and other major entertainment areas. The largest concentration of these large-scale displays can be found in Shibuya, a popular shopping district in Tokyo. Standing outside of the Shibuya JR (Japan Railroad) train station waiting to cross a busy intersection, one is assaulted by a total, at last count, of three large LED displays[4], one at least six stories tall and meshed into a building. All attempt to gain your attention. Here, screens become architecture, screens overtake the physical space, dwarf pedestrians, and launch you into screenspace.

All at once, you find yourself surrounded by things in motion, people with and without keitai, automobiles, images, and type. This experience of screenspace is one that is on a grand scale, but there are countless other examples of screenspace to be experienced in Tokyo. JR Trains now have large hi-definition screens throughout the cars updating us on the weather forecast, news, and latest canned coffee drink as we commute. Plasma screens are also all the rage in Tokyo; having one mounted—hung like a painting—in your apartment makes any image on it a work of art.

Also on the streets of Tokyo, one finds the flickering video arcade. Inside, there is the martial-arts superhero game with its familiar joystick controller along with games having much more developed physical interfaces. One involves a dance floor where users interact with a virtual character and try to keep up with dance beats. Another requires a user to beat on taiko drums as digital images on a screen key you when to do so. These hyper-interactive tangible interfaces mesh screenspace to a physical presence. Like the keitai

Figure 1.

with its individual interaction, these video games are changing the physical and virtual landscape of contemporary Tokyo.

As the video arcades in Tokyo exhibit tangible interfaces fused with screens, the even more common Pachinko parlors now have small-embedded LCDs. The Pachinko parlor itself is a spectacle. Upon entering, the sound of thousands of cascading silver balls all at the control of gravity release a pulsating wave of metal. These machines also play computerized jingles. With the addition of movement from computer graphics on the built-in LCD screens and the drive to gamble, an electrifying effect is created.

As the physical world meshes with the virtual in Tokyo, the screens are friendly and unintimidating. For example, at the JR Shinjuku station, I interface with the ticket machine through its extremely clear GUI as I choose my fare. A 180-yen ticket is presented, a virtual ticket agent bows with a "thank you," and I am on my way. Throughout Tokyo, this interaction with various electronic devices and their screens is polite, apologetic, and clear. Warning beeps or alarms tend to be pleasant and instructive, rather than loud and embarrassing. Cute characters along with strategically placed blinking LEDs explain instructions and directions. It seems that the Japanese are at a much greater ease with technology and its integration into daily life.

This acceptance of technology into daily life in Japan permits screenspace to exist. Major electronic companies such as Sony, Panasonic and Hitachi, to name a few, all release their products for consumer testing on the streets of the Akihabara district in Tokyo. Akihabara—"electric city" in Japanese—is very befitting. On its streets, one finds everything from the latest electronic goods to the raw components that make up these high-tech devices. Here in this bazaar of capacitors, transistors, and micro-controllers, various screens can be found—from simple LED displays to compact high-resolution screens. These raw components found in Akihabara fuel the ingenuity and creativity of Japanese electrical engineers, designers, and artist.

With all of these screens in Tokyo, the keitai, with its mobility and personal space, is the most powerful. The mobility that the keitai exhibits is drastically changing daily life in contemporary Japan and can be compared to some of Tokyo's compact architecture. The keitai is small, functional,

Figure 2.

and mobile. It is necessary for anyone traversing Tokyo's urban landscape. (In fact during my last visit, I found myself feeling awkward having to use the vulnerable payphone.) The ubiquitous capsule hotel also exhibits the same characteristics: functional and necessary in contemporary Tokyo. Positioned near major intersections of travel and entertainment, the capsule hotel allows the businessman[5] who has missed the last train home a sleeping birth one meter by one meter by two meters, sleek yet brutally functional. This functionality and compactness to support all forms of mobility is exemplified in Kisho Kurakabo's Nakagin Capsule Tower found in Ginza. Built in 1970, its 74 capsules are all equipped with a single bed, unit-bath, and entertainment center, initially intended to be purchased by companies to allow temporary stay for the mobile businessman. This compactness and mobility that the Japanese have become accustomed to is essential to the keitai popularity. With the amount of commuting and time spent away from home, the keitai presents itself as a necessary interface for contemporary Japan.

Figure 3.

DIGITAL TYPOGRAPHY: AESTHETIC OF THE PIXEL

With the multitude of screens found in Tokyo, viewing graphics or reading their type can be tiring on the eyes. Type, in particular, can be troublesome: it is often in Japanese and in a vertical format. All images and characters must adhere to various screen resolutions creating, at times, very coarse graphics and jagged type.

On the keitai, one will often find Hiragana, Katakana, and Kanji[6] with little English. Much typography on the keitai is comparable to type on an early Macintosh SE. It is typically black on white, static, and read in the body of an e-mail or a simple interface. It is meant to be personal and clear. This treatment aligns itself with the printed page and Tschichold's New Typography: simple, clear, and clean. But upon closer look, the type is coarse and blocky, adhering to the screen's matrix of pixels.

This coarse and blocky type is not new to typography. Wim Crouwel's 1966 Vomgevers poster for the Stedelijk Museum in Amsterdam used a typeface that adheres to a rigorous grid. Called Stedelijk, this typeface has a very similar look to many "pixel fonts" that are being designed for screens today. Later, in 1967, Crouwel designed New (Neue) Alphabet, a result of his using the first electronic typesetting machines. New Alphabet was a radical design: it removed all diagonal strokes and curves in its individual characters. New Alphabet reduced characters to their essential elements, creating a font that is strictly vertical and horizontal, again adhering to the grid.

More recently, in 1985, Émigré released the Oakland typeface. Designed by Zuzana Licko, it is a coarse bit-mapped family that was a response

Figure 4.

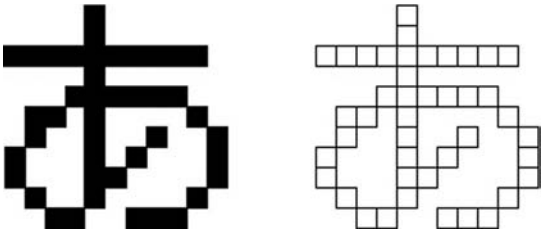


Figure 5.

New Alphabet typeface

to graphic design's shift from tangible to computer screen-based production. This digitally influenced typeface announced the beginnings of the shift in typography from "hands-on" to digital layout.

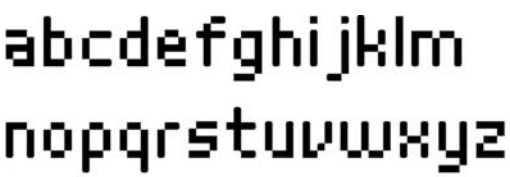
Today, we find ourselves in what I call the aesthetic of the pixel, which in turn is part of screenspace. Currently, there are countless fonts that are specifically designed to be read on the screen. These pixel fonts—with clever names such as Silkscreen, Seven-Net and Superpoint—are designed to be read on very small scales. Their geometry responds to the screen's matrix and, compared to the typeface Oakland, seems much more refined.

This then brings me back to the keitai screen's resolution and pixel dimensions. Most current keitai screen resolutions are around 132x176 pixels, with a few at 360x480 pixels. Color keitai screens range from the basic 256 colors to a crisp 266k colors. The various keitai built-in cameras can capture still images from 176x144 to an astonishing 640x480 pixels and some with even a few seconds of video.

With these achievements in high screen resolutions, KDDI Corp. and Okinawa Cellular released four keitai in August 2002 with built-in cameras that are able to capture up to 15 seconds of smooth running video and audio. Captured video can then be viewed, subtitles and voiceovers can be added, video can be attached to an e-mail and easily sent to another "Movie Keitai" or computer. These keitai take advantage of the third-generation CDMA2000 wireless network which has data transmission speeds of up to 144 kbps allowing captured images to be sent and downloaded smoothly.

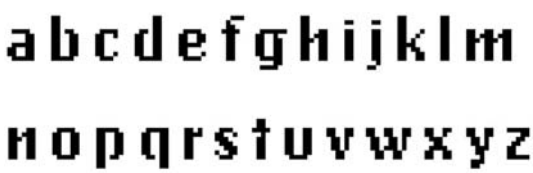
There are hundreds of paid subscription services for the keitai that provide everything from astrological information to financial transactions. All of these services have rich GUI's and some are animated. But probably one of the most interesting uses of graphics and interaction is the map service simply called "Keitai Map" provided by Zenrin. Enter a simple search query on your screen and up comes a vector map image that zooms in and out smoothly. The maps are clear and if used in tandem with one's built-in GPS you will never be lost in Tokyo's labyrinth of streets again.

Figure 5.



Stedelijk typeface

Figure 6.



Oakland typeface

Along with captured images, subscription services, and Japanese characters, an iconic graphic language has emerged for quick messaging on the keitai. Called "emoji," (a shorthand) it evolved from the use of text to make images that would represent a user's emotions. Examples: ^_^ represents the user is happy; ($_$) represents the user is greedy; and (o|o) represents the Japanese superhero Ultra man. There are now hundreds of little emoji 16x16 pixel icons, mostly static, some animated. Emoji has become hugely popular. A typical message—"Meet me for a beer at 7 p.m."—could consist of the emoji for a frosty beer mug and a clock showing 7 p.m. Emoji has become a new typography developed for the keitai.

In contrast to the personal graphics and text created on the keitai, graphics and type on the large LED screens are much more robust and often in full motion with sound, creating the ultimate experience of screenspace. Most screens are typically only a few meters across, but the huge screen called Q Eye, found in Shibuya, is a least six stories tall. Located on the Q Front[7] building in Shibuya, it is an example of architecture becoming a screen or a screen becoming architecture. On its front surface there is a large display made up of three separate screens containing more than 121,600 LED clusters broken into red, green and blue. Q Eye towers over pedestrians as they navigate the intersections around the JR Shibuya station. It is composed of three separate screens: two "banner" screens on the top and bottom and one main screen in the center. Since the screens are synchronized, they create the illusion of one large screen towering six stories tall. Q Eye seems to be always on and in constant motion, displaying videos of Japan's hottest pop stars to advertisements for Sony's new cameras. At night, the glow from its screens, along with the two other small screens in its vicinity, light up the streets, giving Shibuya an eerie feel as colors and images shift and move.

FUTURE TRENDS

As Tokyo's streets shine and flicker from the ever-popular keitai and newly constructed jumbo LED displays, major companies are developing the next generation of screens. Of particular interest for these researchers: improvement of screen resolution, energy usage, thinness, and flexibility. The focus on flexibility and thinness is evident in current research on organic electroluminescence (OEL) and E-ink's screens.

Electronic ink's (E-Ink's) main components consist of thousands of tiny microcapsules that have the diameter of 100 microns (about the width of a human hair). These microcapsules contain both positively and negatively charged particles of white and black color. When a positive electric field is applied to the microcapsule, the white particles become visible; when a negative charge is applied, the black particles become visible. With these microcapsules turning on and off, thousands are necessary to create an image. Arranged in a grid, they function in a similar fashion as a pixel on a computer screen.

These microcapsules are suspended in a liquid, allowing E-Ink to be printed on virtually any surface from plastic to metal. The main advantages that E-Ink demonstrates over current display technologies include low power usage, flexibility, durability, and readability. First introduced in 1999, E-Ink displays draw only 0.1 watts of power and only when changing its display; an E-Ink sign with static information draws no power.

Electronic paper, which is very similar to E-Ink, is being simultaneously developed by Xerox. First developed in the 1970s it functions in a similar manner to E-Ink, but has microscopic balls that are black on one side and white on the other. These balls rotate when an electrical charge is applied, showing either black or white.

E-Ink is currently more desirable because of its simplicity and cost compared to electronic paper, which also requires complicated wiring.

The main drawback with current E-Ink and electronic paper technology is that their screens are only monochromatic (multicolored screens are in development).

Organic electroluminescence displays (OEL) contain thin organic layers made of a carbon-based compound that illuminates brightly when an electric charge is applied. This allows for thin, flexible, cool (temperature), brightly glowing screens that operate on very low power. Monochromatic passive-matrix OEL screens are currently found everywhere from car stereos to keitai. Currently, full color active-matrix OEL displays, driven by thin film transistors (TFT), are limited in size due to production cost.

The future market potential for large OEL screens is tremendous. Major companies (e.g., Pioneer, Sony, Sanyo, Kodak, NEC, and Samsung) have been pouring millions of research dollars into perfecting the production and performance of large full color OEL active-matrix displays. From this research, Sanyo has developed and put into production a 260,000-color OEL screen for use with third-generation multimedia keitai currently found in Japan. This commitment by Sanyo indicates the future direction of keitai screen technology.

The real breakthrough will occur when full color OEL displays reach 10" (diagonal measurement) or more with screen resolutions of 800x600 pixels and greater. In 2001, Sony was the first to produce a prototype screen measuring 13" with a screen resolution of 800x600 pixels and measuring just 1.4 millimeters in thickness. Mass production of these large OEL screens will replace the cathode ray tube and current LCD monitors, bringing us one step closer to jumbo OEL screens.

Pioneer began mass production of monochromatic passive-matrix OEL displays in 1997. Used on Motorola cellphones, these passive-matrix screens do not need integrated circuits or a glass base, allowing the screen to be built on plastic. From this research, Pioneer has prototyped wearable plastic OEL screens attached to winter coats and has begun the mass production of small passive-matrix screens for PDAs.

The cutting edge research that is being done is essential for the future of screenspace. With E-Ink and electronic paper, screens will be more flexible and will require little maintenance and energy consumption. But the real future of screenspace relies upon the research in OEL display technology. With the possibilities of screens being incorporated into clothing and jumbo OEL screens (as well as meshing into newly constructed buildings) every surface has the potential to become a screen.

Along with this vast research being conducted by R&D labs of major companies, two artists—Ryota Kuwakubo and Sugihara Yuki—are expanding our ideas of screens and how we interact with them. Ryota Kuwakubo's work uses low-powered LED screens that are part of highly interactive devices. As users interact with the objects, elegantly designed circuits transmit signals to small clusters of LEDs creating patterns, symbols, and simple smiling faces. Bit-Man, designed to be worn around the neck while dancing, has an LED display of a man that responds to various sensors. The electronic man consists of a grid of red LEDs; the more Bit-Man is shaken, the longer and faster the electronic man dances. If Bit-Man is rotated in any direction, the LED man positions himself upright.

In contemporary Japan, Yuki Sugihara steps away from the traditional use and understanding of screens. She utilizes water as a "screen" to project light and images upon. Sugihara's work, titled "Head- Mounted Water Display," recently exhibited at NTT's Intercommunication Center in Tokyo. It invites viewers to step into a dark room. Inside, a cascading dome of water has been placed in the room's center and users step under the water dome one at a time to view colors and patterns projected upon the fluid surface of the dome. Sugihara brings the natural and the technological together in an elegant display of water, color, and image.

CONCLUSION

As of this writing, Japan leads the world in the development of the small personal keitai screens along with the jumbo LED displays. Times Square in New York City seems to be the only other location that rivals Shibuya with its saturation of screens. Major cellular phone companies in America are now realizing the popularity of wireless web access and have developed limited web access applications[8]. Movies such as *Minority Report* (2002), *The Time Machine* (2002), and *Brazil* (1985) present us with glimpses of what the future could hold for us and how we might interact with screenspace. Screens are becoming more affordable and sleek LCD flat screen monitors are more common, replacing the bulky CRT (cathode ray tube). Also, with research on OEL displays (E-Ink and the next generation of keitai) screenspace and cellspace will eventually become one.

As these technologies mature and new technologies are discovered, displays will not only get clearer, but will have the ability to be anywhere and everywhere. Screenspace and cellspace will become one, and urban landscapes will flicker brightly in the night sky.

REFERENCES

Barlow, J., & Resnic, D. (June 6, 2002). *E Ink unveils world's thinnest active matrix display.* Retrieved from: http://www.eink.com/news/releases/pr60.html.

Bennahum, D S. (n.d.). *CellSpace*. Retrieved from: http://www.memex.org/meme4-03.html; http://whatis.techtarget.com/definition/0,,sid9_gci211762,00.html.

Grotz, V. (1998). *Color & type for the screen crans.* RotoVision SA.

Miyake, K. (October 7, 2001). *Organic EL displays make many appearances.* Retrieved from: http://www.cnn.com/2001/TECH/ptech/10/07/organic.el.idg/.

Nelson, T. & Revie, J. (1996). *Thin-film transistor LCD displays.* Retrieved from: http://www.cs.ndsu.nodak.edu/~revie/amlcd/.

Ross, M.F. (1978). *Beyond Metabolism: The New Japanese Architecture*. New York: Architectural Record.

Suzuki, A. (2001). *Do Android Crows Fly Over the Skies of an Electronic Tokyo?* (Trans. J. Keith Vincent). London: Architectural Association.

Wurster, C. (2002). *Computers: An Illustrated History.* Koln, Taschen.

ENDNOTES

1. Hertzian Space is used by Anthony Dunne and Fiona Raby to described the invisible spectrum of electromagnetic waves that are emitted from various devices.
2. Japanese for cellphone and translates to "portable".
3. Sha is short for photograph in Japanese.
4. As of July 2002, there were three large LED displays at Shibuya crossing.
5. Capsule hotels are typically for men only.
6. Hirigana and Katakana are phonetic characters where Kanji is more complex and each character represents various words depending on context.
7. Q eye is a LED screen located on the front of the Q front building at Shibuya crossing in Tokyo, Japan.
8. mMode from AT&T Wireless and PCS Vision from Sprint are bringing high-speed access to their cellular phone users. These

services allow web access, gaming, instant messaging, custom ring tones and images to be sent from on phone to the other.

This work was previously published in Computer Graphics and Multimedia: Applications, Problems and Solutions, edited by J. DiMarco, pp. 40-53, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.11
Audio Watermarking:
Properties, Techniques and Evaluation

Andrés Garay Acevedo
Georgetown University, USA

ABSTRACT

The recent explosion of the Internet as a collaborative medium has opened the door for people who want to share their work. Nonetheless, the advantages of such an open medium can pose very serious problems for authors who do not want their works to be distributed without their consent. As new methods for copyright protection are devised, expectations around them are formed and sometimes improvable claims are made. This chapter covers one such technology: audio watermarking. First, the field is introduced, and its properties and applications are discussed. Then, the most common techniques for audio watermarking are reviewed, and the framework is set for the objective measurement of such techniques. The last part of the chapter proposes a novel test and a set of metrics for thorough benchmarking of audio watermarking schemes. The development of such a benchmark constitutes a first step towards the standardization of the requirements and properties that such systems should display.

INTRODUCTION

The recent explosion of the Internet as a collaborative medium has opened the door for people who want to share their work. Nonetheless, the advantages of such an open medium can pose very serious problems for authors who do not want their works to be distributed without their consent. The digital nature of the information that traverses through modern networks calls for new and improved methods for copyright protection[1].

In particular, the music industry is facing several challenges (as well as opportunities) as it tries to adapt its business to the new medium. Content protection is a key factor towards a comprehensive information commerce infrastructure (Yeung, 1998), and the industry expects new technologies will help them protect against the misappropriation of musical content.

One such technology, digital watermarking, has recently brought a tide of publicity and controversy. It is an emerging discipline, derived from an older science: steganography, or the hiding of a secret message within a seemingly

innocuous cover message. In fact, some authors treat watermarking and steganography as equal concepts, differentiated only by their final purpose (Johnson, Duric, & Jajodia, 2001).

As techniques for digital watermarking are developed, claims about their performance are made public. However, different metrics are typically used to measure performance, making it difficult to compare both techniques and claims. Indeed, there are no standard metrics for measuring the performance of watermarks for digital audio. Robustness does not correspond to the same criteria among developers (Kutter & Petitcolas, 1999). Such metrics are needed before we can expect to see a commercial application of audio watermarking products with a provable performance.

The objective of this chapter is to propose a methodology, including performance metrics, for evaluating and comparing the performance of digital audio watermarking schemes. In order to do this, it is necessary first to provide a clear definition of what constitutes a watermark and a watermarking system in the context of digital audio. This is the topic of the second section, which will prove valuable later in the chapter, as it sets a framework for the development of the proposed test.

After a clear definition of a digital watermark has been presented, a set of key properties and applications of digital watermarks can be defined and discussed. This is done in the third section, along with a classification of audio watermarking schemes according to the properties presented. The importance of these properties will be reflected on the proposed tests, discussed later in the chapter. The survey of different applications of watermarking techniques gives a practical view of how the technology can be used in a commercial and legal environment. The specific application of the watermarking scheme will also determine the actual test to be performed to the system.

The fourth section presents a survey of specific audio watermarking techniques developed. Five general approaches are described: amplitude modification, dither watermarking, echo watermarking, phase distortion, and spread spectrum watermarking. Specific implementations of watermarking algorithms (i.e., test subjects) will be evaluated in terms of these categories[2].

The next three sections describe how to evaluate audio watermarking technologies based on three different parameters: fidelity, robustness, and imperceptibility. Each one of these parameters will be precisely defined and discussed in its respective section, as they directly reflect the interests of the three main actors involved in the communication process[3]: sender, attacker, and receiver, respectively.

Finally, the last section provides an account on how to combine the three parameters described above into a single performance measure of quality. It must be stated, however, that this measure should be dependant upon the desired application of the watermarking algorithm (Petitcolas, 2000).

The topics discussed in this chapter come not only from printed sources but also from very productive discussions with some of the active researchers in the field. These discussions have been conducted via e-mail, and constitute a rich complement to the still low number of printed sources about this topic. Even though the annual number of papers published on watermarking has been nearly doubling every year in the last years (Cox, Miller, & Bloom, 2002), it is still low. Thus it was necessary to augment the literature review with personal interviews.

WATERMARKING: A DEFINITION

Different definitions have been given for the term *watermarking* in the context of digital content. However, a very general definition is given by Cox et al. (2002), which can be seen as application independent: "We define watermarking as the practice of imperceptibly altering a Work to

embed a message about that Work". In this definition, the word *work* refers to a specific song, video or picture[4].

A crucial point is inferred by this definition, namely that the information hidden within the work, the watermark itself, contains information about the work where it is embedded. This characteristic sets a basic requirement for a watermarking system that makes it different from a general steganographic tool. Moreover, by distinguishing between embedded data that relate to the cover work and hidden data that do not, we can derive some of the applications and requirements of the specific method. This is exactly what will be done later.

Another difference that is made between watermarking and steganography is that the former has the additional notion of robustness against attacks (Kutter & Hartung, 2000). This fact also has some implications that will be covered later on.

Finally, if we apply Cox's definition of watermarking into the field of audio signal processing, a more precise definition, this time for audio watermarking, can be stated. Digital audio watermarking is defined as the process of "embedding a user specified bitsream in digital audio such that the addition of the watermark (bitstream) is perceptually insignificant" (Czerwinski, Fromm, & Hodes, 1999).

This definition should be complemented with the previous one, so that we do not forget the watermark information refers to the digital audio file.

Elements of an Audio Watermarking System

Embedded watermarks are recovered by running the inverse process that was used to embed them in the cover work, that is, the original work. This means that all watermarking systems consist of at least two generic building blocks: a watermark embedding system and a watermark recovery system.

Figure 1 shows a basic watermarking scheme, in which a watermark is both embedded and recovered in an audio file. As can be seen, this process might also involve the use of a secret key. In general terms, given the audio file A, the watermark W and the key K, the embedding process is a mapping of the form $A \times K \times W \rightarrow A'$

Conversely, the recovery or extraction process receives a tentatively watermarked audio file A', and a recovery key K' (which might be equal to K), and it outputs either the watermark W or a confidence measure about the existence of W (Petitcolas, Anderson, & G., 1999).

At this point it is useful to attempt a formal definition of a watermarking system, based on that of Katzenbeisser (2000), and which takes into account the architecture of the system. The quintuple $\xi = \langle C, W, K, D_k, E_k \rangle$, where C is the set of possible audio *covers*[5], W the set of *watermarks* with $|C| \geq |W|$, K the set of secret *keys*, $E_k: C \times K \times W \rightarrow C$ the embedding function and $D_k: C \times K \rightarrow W$ the extraction function, with the property that $D_k(E_k(c, k, w)\, k) = w$ for all $w \in W$, $c \in C$ and $k \in K$ is called a *secure audio watermarking system*.

Figure 1. Basic watermarking system

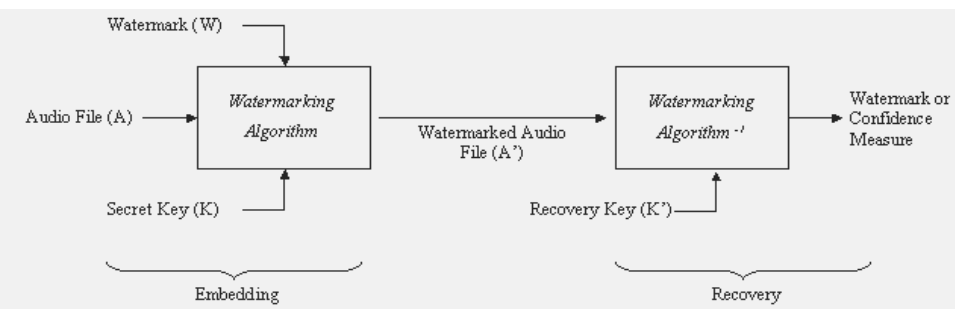
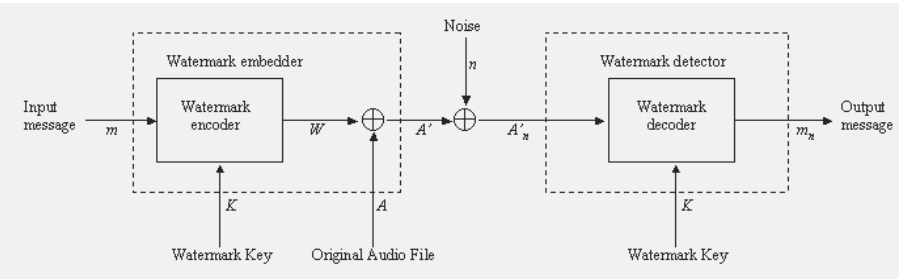


Figure 2. Watermark communication process



This definition is almost complete, but it fails to cover some special cases. Some differences might arise between a real world system, and the one just defined; for example, some detectors may not output the watermark *W* directly but rather report the existence of it. Nonetheless, it constitutes a good approximation towards a widely accepted definition of an audio watermarking system.

If one takes into account the small changes that a marking scheme can have, a detailed classification of watermarking schemes is possible. In this classification, the different schemes fall into three categories, depending on the set of inputs and outputs (Kutter & Hartung, 2000). Furthermore, a specific and formal definition for each scheme can be easily given by adapting the definition just given for an audio watermarking system.

Private watermarking systems require the original audio *A* file in order to attempt recovery of the watermark *W*. They may also require a copy of the embedded watermark and just yield a yes or no answer to the question: does *A′* contain *W*?

Semi-private watermarking schemes do not use the original audio file for detection, but they also answer the yes/no question shown above. This could be described by the relation $A' \times K \times W \rightarrow$ {"Yes", "No"}.

Public watermarking (also known as *blind* or *oblivious* watermarking) requires neither the original file *A*, nor the embedded watermark *W*. These systems just extract *n* bits of information from the watermarked audio file. As can be seen, if a key is used then this corresponds to the definition given for a *secure watermarking system*.

Watermark as a Communication Process

A watermarking process can be modeled as a communication process. In fact, this assumption is used throughout this chapter. This will prove to be beneficial in the next chapter when we differentiate between the requirements of the content owner and consumer. A more detailed description of this model can be found in Cox et al. (2002).

In this framework, the watermarking process is viewed as a transmission channel through which the watermark message is communicated. Here the cover work is just part of the channel. This is depicted in Figure 2, based on that from Cox et al. (2002).

In general terms, the embedding process consists of two steps. First, the watermark message *m* is mapped into an added pattern[6] W_a, of the same type and dimension as the cover work *A*. When watermarking audio, the watermark encoder produces an audio signal. This mapping may be done with a watermark key *K*. Next, W_a is embedded into the cover work in order to produce the watermarked audio file *A′*.

After the pattern is embedded, the audio file is processed in some way. This is modeled as the addition of noise to the signal, which yields a noisy work A'_n. The types of processing performed on

the work will be discussed later, as they are of no importance at this moment. However, it is important to state the presence of noise, as any transmission medium will certainly induce it.

The watermark detector performs a process that is dependant on the type of watermarking scheme. If the decoder is a *blind or public decoder*, then the original audio file *A* is not needed during the recovery process, and only the key *K* is used in order to decode a watermark message m_n. This is the case depicted in Figure 2, as it is the one of most interest to us.

Another possibility is for the detector to be *informed*. In this case, the original audio cover *A* must be extracted from A'_n in order to yield W_n, prior to running the decoding process. In addition, a confidence measure can be the output of the system, rather than the watermark message.

PROPERTIES, CLASSIFICATION AND APPLICATIONS

After a proper definition of a watermarking scheme, it is possible now to take a look at the fundamental properties that comprise a watermark. It can be stated that an ideal watermarking scheme will present all of the characteristics here detailed, and this ideal type will be useful for developing a quality test.

However, in practice there exists a fundamental trade-off that restricts watermark designers. This fundamental trade-off exists between three key variables: robustness, payload and perceptibility (Cox, Miller, Linnartz, & Kalker, 1999; Czerwinski et al., 1999; Johnson et al., 2001; Kutter & Petitcolas, 1999; Zhao, Koch, & Luo, 1998). The relative importance given to each of these variables in a watermarking implementation depends on the desired application of the system.

Fundamental Properties

A review of the literature quickly points out the properties that an ideal watermarking scheme should possess (Arnold, 2000; Boney, Tewfik & Hamdy, 1996; Cox, Miller, & Bloom, 2000; Cox et al., 1999, 2002; Kutter & Hartung, 2000; Kutter & Petitcolas, 1999; Swanson, Zhu, Tewfik, & Boney, 1998). These are now discussed.

Imperceptibility. "The watermark should not be noticeable … nor should [it] degrade the quality of the content" (Cox et al., 1999). In general, the term refers to a similarity between the original and watermarked versions of the cover work.

In the case of audio, the term *audibility* would be more appropriate; however, this could create some confusion, as the majority of the literature uses *perceptibility*. This is the same reason why the term *fidelity* is not used at this point, even though Cox et al. (1999) point out that if a watermark is truly imperceptible, then it can be removed by perceptually-based lossy compression algorithms. In fact, this statement will prove to be a problem later when trying to design a measure of watermark perceptibility. Cox's statement implies that some sort of perceptibility criterion must be used not only to design the watermark, but to quantify the distortion as well. Moreover, it implies that this distortion must be measured at the point where the audio file is being presented to the consumer/receiver.

If the distortion is measured at the receiver's end, it should also be measured at the sender's. That is, the distortion induced by a watermark must also be measured before any transmission process. We will refer to this characteristic at the sending end by using the term *fidelity.*

This distinction between the terms *fidelity* and *imperceptibility* is not common in the literature, but will be beneficial at a later stage. Differentiating between the amount and characteristics of the noise or distortion that a watermark introduces in a signal before and after the transmission process takes into account the different expectations that content owners and consumers have from the technology. However, this also implies that the metric used to evaluate this effect must be different at these points. This is exactly what will be done later on this chapter.

Artifacts introduced through a watermarking process are not only annoying and undesirable, but may also reduce or destroy the commercial value of the watermarked data (Kutter & Hartung, 2000). Nonetheless, the perceptibility of the watermark can increase when certain operations are performed on the cover signal.

Robustness refers to the ability to detect the watermark after common signal processing operations and hostile attacks. Examples of common operations performed on audio files include noise reduction, volume adjustment or normalization, digital to analog conversion, and so forth. On the other hand, a hostile attack is a process specifically designed to remove the watermark.

Not all watermarking applications require robustness against all possible signal processing operations. Only those operations likely to occur between the embedding of the mark and the decoding of it should be addressed. However, the number and complexity of attack techniques is increasing (Pereira, Voloshynovskiy, Madueño, Marchand-Maillet, & Pun, 2001; Voloshynovskiy, Pereira, Pun, Eggers, & Su, 2001), which means that more scenarios have to be taken into account when designing a system. A more detailed description of these attacks is given in the sixth section.

Robustness deals with two different issues; namely the presence and detection of the watermark after some processing operation. It is not necessary to remove a watermark to render it useless; if the detector cannot report the presence of the mark then the attack can be considered successful. This means that a watermarking scheme is robust when it is able to withstand a series of attacks that try to degrade the quality of the embedded watermark, up to the point where it's removed, or its recovery process is unsuccessful. "No such perfect method has been proposed so far, and it is not clear yet whether an absolutely secure watermarking method exists at all" (Kutter & Hartung, 2000).

Some authors prefer to talk about *tamper resistance* or even *security* when referring to hostile attacks; however, most of the literature encompasses this case under the term *robustness.*

The *effectiveness* of a watermarking system refers to the probability that the output of the embedder will be watermarked. In other words, it is the probability that a watermark detector will recognize the watermark immediately after inserting it in the cover work. What is most amazing about this definition is the implication that a watermarking system might have an effectiveness of less than 100%. That is, it is possible for a system to generate marks that are not fully recoverable even if no processing is done to the cover signal. This happens because perfect effectiveness comes at a very high cost with respect to other properties, such as perceptibility (Cox et al., 2002). When a known watermark is not successfully recovered by a detector it is said that a false negative, or type-II error, has occurred (Katzenbeisser, 2000).

Depending on the application, one might be willing to sacrifice some performance in exchange for other characteristics. For example, if extremely high fidelity is to be achieved, one might not be able to successfully watermark certain type of works without generating some kind of distortion. In some cases, the effectiveness can be determined analytically, but most of the time it has to be estimated by embedding a large set of works with a given watermark and then trying to extract that mark. However, the statistical characteristics of the test set must be similar to those of the works that will be marked in the real world using the algorithm.

Data payload. In audio watermarking this term refers to the number of embedded bits per second that are transmitted. A watermark that encodes *N* bits is referred to as an N-bit watermark, and can be used to embed 2^N different messages. It must be said that there is a difference between the encoded message *m*, and the actual bitstream that is embedded in the audio cover work. The

latter is normally referred to as a *pseudorandom* (PN) *sequence*.

Many systems have been proposed where only one possible watermark can be embedded. The detector then just determines whether the watermark is present or not. These systems are referred to as *one-bit watermarks*, as only two different values can be encoded inside the watermark message. In discussing the data payload of a watermarking method, it is also important to distinguish between the number of distinct watermarks that may be inserted, and the number of watermarks that may be detected by a single iteration with a given watermark detector. In many watermarking applications, each detector need not test for all the watermarks that might possibly be present (Cox et al., 1999). For example, one might insert two different watermarks into the same audio file, but only be interested in recovering the last one to be embedded.

Other Properties

Some of the properties reviewed in the literature are not crucial for testing purposes; however they must be mentioned in order to make a thorough description of watermarking systems.

- **False positive rate.** A false positive or type-I error is the detection of a watermark in a work that does not actually contain one. Thus a false positive rate is the expected number of false positives in a given number of runs of the watermark detector. Equivalently, one can detect the probability that a false positive will occur in a given detector run.
  In some applications a false positive can be catastrophic. For example, imagine a DVD player that incorrectly determines that a legal copy of a disk (for example a homemade movie) is a non-factory-recorded disk and refuses to play it. If such an error is common, then the reputation of DVD players and consequently their market can be seriously damaged.
- **Statistical invisibility.** This is needed in order to prevent unauthorized detection and/or removal. Performing statistical tests on a set of watermarked files should not reveal any information about the nature of the embedded information, nor about the technique used for watermarking (Swanson et al., 1998). Johnson et al. (2001) provide a detailed description of known signatures that are created by popular information hiding tools. Their techniques can be also extended for use in some watermarking systems.
- **Redundancy.** To ensure robustness, the watermark information is embedded in multiple places on the audio file. This means that the watermark can usually be recovered from just a small portion of the watermarked file.
- **Compression ratio,** or similar compression characteristics as the original file. Audio files are usually compressed using different schemes, such as MPEG-Layer 3 audio compression. An audio file with an embedded watermark should yield a similar compression ratio as its unmarked counterpart, so that its value is not degraded. Moreover, the compression process should not remove the watermark.
- **Multiple watermarks.** Multiple users should be able to embed a watermark into an audio file. This means that a user has to ideally be able to embed a watermark without destroying any preexisting ones that might be already residing in the file. This must hold true even if the watermarking algorithms are different.
- **Secret keys.** In general, watermarking systems should use one or more cryptographically secure keys to ensure that the watermark cannot be manipulated or erased. This is important because once a watermark can be read by someone, this same person might alter it since both the location and embedding algorithm of the mark will be known (Kutter & Hartung, 2000). It is not

safe to assume that the embedding algorithm is unknown to the attacker.

As the security of the watermarking system relies in part on the use of secret keys, the *keyspace* must be large, so that a brute force attack is impractical. In most watermarking systems the key is the PN-pattern itself, or at least is used as a seed in order to create it. Moreover, the watermark message is usually encrypted first using a cipher key, before it is embedded using the watermark key. This practice adds security at two different levels. In the highest level of secrecy, the user cannot read or decode the watermark, or even detect its presence. The second level of secrecy permits any user to detect the presence of the watermark, but the data cannot be decoded without the proper key. Watermarking systems in which the key is known to various detectors are referred to as unrestricted-key watermarks. Thus, algorithms for use as unrestricted-key systems must employ the same key for every piece of data (Cox et al., 1999). Those systems that use a different key for each watermark (and thus the key is shared by only a few detectors) are known as restricted-key watermarks.

- **Computational cost.** The time that it takes for a watermark to be embedded and detected can be a crucial factor in a watermarking system. Some applications, such as broadcast monitoring, require real time watermark processing and thus delays are not acceptable under any circumstances. On the other hand, for court disputes (which are rare), a detection algorithm that takes hours is perfectly acceptable as long as the effectiveness is high.

Additionally, the number of embedders and detectors varies according to the application. This fact will have an effect on the *cost* of the watermarking system. Applications such as DVD copy control need few embedders but a detector on each DVD player; thus the cost of recovering should be very low, while that of embedding could be a little higher[7]. Whether the algorithms are implemented as plug-ins or dedicated hardware will also affect the economics of deploying a system.

Different Types of Watermarks

Even though this chapter does not relate to all kinds of watermarks that will be defined, it is important to state their existence in order to later derive some of the possible applications of watermarking systems.

- **Robust watermarks** are simply watermarks that are robust against attacks. Even if the existence of the watermark is known, it should be difficult for an attacker to destroy the embedded information without the knowledge of the key8. An implication of this fact is that the amount of data that can be embedded (also known as the payload) is usually smaller than in the case of steganographic methods. It is important to say that watermarking and steganographic methods are more complementary than competitive.
- **Fragile watermarks** are marks that have only very limited robustness (Kutter & Hartung, 2000). They are used to detect modifications of the cover data, rather than convey inerasable information, and usually become invalid after the slightest modification of a work. Fragility can be an advantage for authentication purposes. If a very fragile mark is detected intact in a work, we can infer that the work has probably not been altered since the watermark was embedded (Cox et al., 2002). Furthermore, even semi-fragile watermarks can help localize the exact location where the tampering of the cover work occurred.
- **Perceptible watermarks,** as the name states, are those that are easily perceived by

the user. Although they are usually applied to images (as visual patterns or logos), it is not uncommon to have an audible signal overlaid on top of a musical work, in order to discourage illegal copying. As an example, the IBM Digital Libraries project (Memon & Wong, 1998; Mintzer, Magerlein, & Braudaway, 1996) has developed a visible watermark that modifies the brightness of an image based on the watermark data and a secret key. Even though perceptible watermarks are important for some special applications, the rest of this chapter focuses on imperceptible watermarks, as they are the most common.

- **Bitstream watermarks** are marks embedded directly into compressed audio (or video) material. This can be advantageous in environments where compressed bitstreams are stored in order to save disk space, like Internet music providers.
- **Fingerprinting** and labeling denote special applications of watermarks. They relate to watermarking applications where information such as the creator or recipient of the data is used to form the watermark. In the case of fingerprinting, this information consists of a unique code that uniquely identifies the recipient, and that can help to locate the source of a leak in confidential information. In the case of labeling, the information embedded is a unique data identifier, of interest for purposes such as library retrieving. A more thorough discussion is presented in the next section.

Watermark Applications

In this section the seven most common application for watermarking systems are presented. What is more important, all of them relate to the field of audio watermarking. It must be kept in mind that each of these applications will require different priorities regarding the watermark's properties that have just been reviewed.

- **Broadcast monitoring.** Different individuals are interested in broadcast verification. Advertisers want to be sure that the ads they pay for are being transmitted; musicians want to ensure that they receive royalty payments for the air time spent on their works.

  While one can think about putting human observers to record what they see or hear on a broadcast, this method becomes costly and error prone. Thus it is desirable to replace it with an automated version, and digital watermarks can provide a solution. By embedding a unique identifier for each work, one can monitor the broadcast signal searching for the embedded mark and thus compute the air time. Other solutions can be designed, but watermarking has the advantage of being compatible with the installed broadcast equipment, since the mark is included within the signal and does not occupy extra resources such as other frequencies or header files. Nevertheless, it is harder to embed a mark than to put it on an extra header, and content quality degradation can be a concern.
- **Copyright owner identification.** Under U.S. law, the creator of an original work holds copyright to it the instant the work is recorded in some physical form (Cox et al., 2002). Even though it is not necessary to place a copyright notice in distributed copies of work, it is considered a good practice, since a court can award more damages to the owner in the case of a dispute.

  However, textual copyright notices[9] are easy to remove, even without intention. For example, an image may be cropped prior to publishing. In the case of digital audio the problem is even worse, as the copyright notice is not visible at all times.

  Watermarks are ideal for including copyright notices into works, as they can be both imperceptible and inseparable from the cover

that contains them (Mintzer, Braudaway, & Bell, 1998). This is probably the reason why copyright protection is the most prominent application of watermarking today (Kutter & Hartung, 2000). The watermarks are used to resolve rightful ownership, and thus require a very high level of robustness (Arnold, 2000). Furthermore, additional issues must be considered; for example, the marks must be unambiguous, as other parties can try to embed counterfeit copyright notices. Nonetheless, it must be stated that the legal impact of watermark copyright notices has not yet been tested in court.

- **Proof of ownership.** Multimedia owners may want to use watermarks not just to identify copyright ownership, but also to actually prove ownership. This is something that a textual notice cannot easily do, since it can be forged.

  One way to resolve an ownership dispute is by using a central repository, where the author registers the work prior to distribution. However, this can be too costly[10] for many content creators. Moreover, there might be lack of evidence (such as sketch or film negatives) to be presented at court, or such evidence can even be fabricated.

  Watermarks can provide a way for authenticating ownership of a work. However, to achieve the level of security required for proof of ownership, it is probably necessary to restrict the availability of the watermark detector (Cox et al., 2002). This is thus not a trivial task.
- **Content authentication.** In authentication applications the objective is to detect modifications of the data (Arnold, 2000). This can be achieved with fragile watermarks that have low robustness to certain modifications. This proves to be very useful, as it is becoming easier to tamper with digital works in ways that are difficult to detect by a human observer.

  The problem of authenticating messages has been well studied in cryptography; however, watermarks are a powerful alternative as the signature is embedded directly into the work. This eliminates the problem of making sure the signature stays with the work. Nevertheless, the act of embedding the watermark must not change the work enough to make it appear invalid when compared with the signature. This can be accomplished by separating the cover work in two parts: one for which the signature is computed, and the other where it is embedded.

  Another advantage of watermarks is that they are modified along with the work. This means that in certain cases the location and nature of the processing within the audio cover can be determined and thus inverted. For example, one could determine if a lossy compression algorithm has been applied to an audio file[11].
- **Transactional watermarks.** This is an application where the objective is to convey information about the legal recipient of digital data, rather than the source of it. This is done mainly to identify single distributed copies of data, and thus monitor or trace back illegally produced copies of data that may circulate12.

  The idea is to embed a unique watermark in each distributed copy of a work, in the process we have defined as fingerprinting. In these systems, the watermarks must be secure against a collusion attack, which is explained in the sixth section, and sometimes have to be extracted easily, as in the case of automatic Web crawlers that search for pirated copies of works.
- **Copy control/device control.** Transactional watermarks as well as watermarks for monitoring, identification, and proof of ownership do not prevent illegal copying (Cox et al., 2000). Copy protection is difficult to achieve in open systems, but might be desirable in

proprietary ones. In such systems it is possible to use watermarks to indicate if the data can be copied or not (Mintzer et al., 1998).

The first and strongest line of defense against illegal copying is encryption, as only those who possess the decryption key can access the content. With watermarking, one could do a very different process: allow the media to be perceived, yet still prevent it from being recorded. If this is the case, a watermark detector must be included on every manufactured recorder, preferably in a tamper resistant device. This constitutes a serious nontechnical problem, as there is no natural incentive for recording equipment manufacturers to include such a detector on their machines. This is due to the fact that the value of the recorder is reduced from the point of view of the consumer.

Similarly, one could implement **play control,** so that illegal copies can be made but not played back by compliant equipment. This can be done by checking a media signature, or if the work is properly encrypted for example. By mixing these two concepts, a buyer will be left facing two possibilities: buying a compliant device that cannot play pirated content, or a noncompliant one that can play pirated works but not legal ones.

In a similar way, one could control a playback device by using embedded information in the media they reproduce. This is known as device control. For example, one could signal how a digital audio stream should be equalized, or even extra information about the artist. A more extreme case can be to send information in order to update the firmware of the playback device while it is playing content, or to order it to shut down at a certain time. This method is practical, as the need for a signaling channel can be eliminated.

- **Covert communication.** Even though it contradicts the definition of watermark given before, some people may use watermarking systems in order to hide data and communicate secretly. This is actually the realm of steganography rather than watermarking, but many times the boundaries between these two disciplines have been blurred. Nonetheless, in the context of this chapter, the hidden message is not a watermark but rather a robust covert communication.

  The use of watermarks for hidden annotation (Zhao et al., 1998), or labeling, constitutes a different case, where watermarks are used to create hidden labels and annotations in content such as medical imagery or geographic maps, and indexes in multimedia content for retrieval purposes. In these cases, the watermark requirements are specific to the actual media where the watermark will be embedded. Using a watermark that distorts a patient's radiography can have serious legal consequences, while the recovery speed is crucial in multimedia retrieval.

AUDIO WATERMARKING TECHNIQUES

In this section the five most popular techniques for digital audio watermarking are reviewed. Specifically, the different techniques correspond to the methods for merging (or inserting) the cover data and the watermark pattern into a single signal, as was outlined in the communication model of the second section.

There are two critical parameters to most digital audio representations: sample quantization method and temporal sampling rate. Data hiding in audio signals is especially challenging, because the human auditory system (HAS) operates over a wide dynamic range. Sensitivity to additive random noise is acute. However, there are some "holes" available. While the HAS has a large

dynamic range, it has a fairly small differential range (Bender, Gruhl, Morimoto, & Lu, 1996). As a result, loud sounds tend to mask out quiet sounds. This effect is known as *masking*, and will be fully exploited in some of the techniques presented here (Swanson et al., 1998).

These techniques do not correspond to the actual implementation of commercial products that are available, but rather constitute the basis for some of them. Moreover, most real world applications can be considered a particular case of the general methods described below.

Finally, it must be stated that the methods explained are specific to the domain of audio watermarking. Several other techniques that are very popular for hiding marks in other types of media, such as discrete cosine transform (DCT) coefficient quantization in the case of digital images, are not discussed. This is done because the test described in the following sections is related only to watermarking of digital audio.

Amplitude Modification

This method, also known as *least significant bit (LSB) substitution*, is both common and easy to apply in both steganography and watermarking (Johnson & Katzenbeisser, 2000), as it takes advantage of the quantization error that usually derives from the task of digitizing the audio signal.

As the name states, the information is encoded into the least significant bits of the audio data. There are two basic ways of doing this: the lower order bits of the digital audio signal can be fully substituted with a pseudorandom (PN) sequence that contains the watermark message *m*, or the PN-sequence can be embedded into the lower order bitstream using the output of a function that generates the sequence based on both the n^{th} bit of the watermark message and the n^{th} sample of the audio file (Bassia & Pitas, 1998; Dugelay & Roche, 2000).

Ideally, the embedding capacity of an audio file with this method is 1 kbps per 1 kHz of sampled data. That is, if a file is sampled at 44 kHz then it is possible to embed 44 kilobits on each second of audio. In return for this large channel capacity, audible noise is introduced. The impact of this noise is a direct function of the content of the host signal. For example, crowd noise during a rock concert would mask some of the noise that would be audible in a string quartet performance. Adaptive data attenuation has been used to compensate for this variation in content (Bender et al., 1996). Another option is to shape the PN-sequence itself so that it matches the audio masking characteristics of the cover signal (Czerwinski et al., 1999).

The major disadvantage of this method is its poor immunity to manipulation. Encoded information can be destroyed by channel noise, resampling, and so forth, unless it is encoded using redundancy techniques. In order to be robust, these techniques reduce the data rate, often by one to two orders of magnitude. Furthermore, in order to make the watermark more robust against localized filtering, a pseudorandom number generator can be used to spread the message over the cover in a random manner. Thus, the distance between two embedded bits is determined by a secret key (Johnson & Katzenbeisser, 2000). Finally, in some implementations the PN-sequence is used to retrieve the watermark from the audio file. In this way, the watermark acts at the same time as the key to the system.

Recently proposed systems use amplitude modification techniques in a transform space rather than in the time (or spatial) domain. That is, a transformation is applied to the signal, and then the least significant bits of the coefficients representing the audio signal *A* on the transform domain are modified in order to embed the watermark *W*. After the embedding, the inverse transformation is performed in order to obtain the watermarked audio file *A'*. In this case, the technique is also known as *coefficient quantization*. Some of the transformations used for watermarking are the

discrete Fourier transform (DFT), discrete cosine transform (DCT), Mellin-Fourier transform, and wavelet transform (Dugelay & Roche, 2000). However, their use is more popular in the field of image and video watermarking.

Dither Watermarking

Dither is a noise signal that is added to the input audio signal to provide better sampling of that input when digitizing the signal (Czerwinski et al., 1999). As a result, distortion is practically eliminated, at the cost of an increased noise floor.

To implement dithering, a noise signal is added to the input audio signal with a known probability distribution, such as Gaussian or triangular. In the particular case of dithering for watermark embedding, the watermark is used to modulate the dither signal. The host signal (or original audio file) is quantized using an associated dither quantizer (RLE, 1999). This technique is known as *quantization index modulation* (QIM) (Chen & Wornell, 2000).

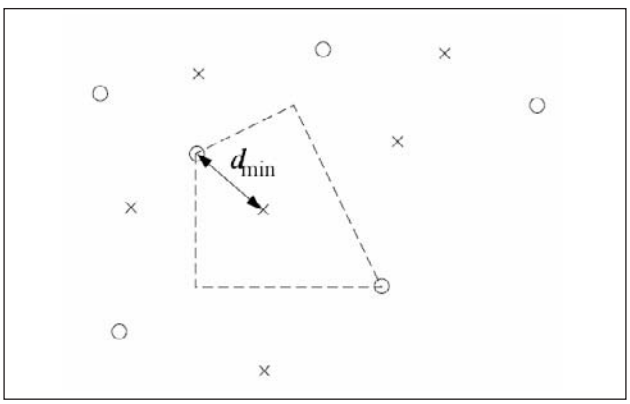
For example, if one wishes to embed one bit (m=1 or m=2) in the host audio signal A then one would use two different quantizers, each one representing a possible value for m. If the two quantizers are shifted versions of each other, then they are called *dither quantizers,* and the process is that of *dither modulation.* Thus, QIM refers to embedding information by first modulating an index or sequence of indices with the embedded information and then quantizing the host signal with the associated quantizer or sequence of quantizers (Chen & Wornell, 1999).

A graphical view of this technique is shown in Figure 3, taken from Chen (2000). Here, the points marked with X's and O's belong to two different quantizers, each with an associated index; that is, each one embedding a different value. The distance d_{min} can be used as an informal measure of robustness, while the size of the quantization cells (one is shown in the figure) measures the distortion on the audio file. If the watermark message m=1, then the audio signal is quantized to the nearest X. If m=2 then it is quantized to the nearest O.

The two quantizers must not intersect, as can be seen in the figure. Furthermore, they have a discontinuous nature. If one moves from the interior of the cell to its exterior, then the corresponding value of the quantization function jumps from an X in the cell's interior to one X on its exterior. Finally, as noted above, the number of quantizers in the ensemble determines the information-embedding rate (Chen & Wornell, 2000).

As was said above, in the case of dither modulation, the quantization cells of any quantizer in the ensemble are shifted versions of the cells of any other quantizer being used as well. The

Figure 3. A graphical view of the QIM technique



shifts traditionally correspond to pseudorandom vectors called the *dither* vectors. For the task of watermarking, these vectors are modulated with the watermark, which means that each possible embedded signal maps uniquely to a different dither vector. The host signal A is then quantized with the resulting dithered quantizer in order to crate the watermarked audio signal A'.

Echo Watermarking

Echo watermarking attempts to embed information on the original discrete audio signal $A(t)$ by introducing a repeated version of a component of the audio signal with small enough offset (or delay), initial amplitude and decay rate $\alpha A(t - \Delta t)$ to make it imperceptible. The resulting signal can be then expressed as $A'(t) = A(t) + \alpha A(t - \Delta t)$.

In the most basic echo watermarking scheme, the information is encoded in the signal by modifying the delay between the signal and the echo. This means that two different values Δt and $\Delta t'$ are used in order to encode either a zero or a one. Both offset values have to be carefully chosen in a way that makes the watermark both inaudible and recoverable (Johnson & Katzenbeisser, 2000).

As the offset between the original and the echo decreases, the two signals blend. At a certain point, the human ear cannot distinguish between the two signals. The echo is perceived as added resonance (Bender et al., 1996). This point is hard to determine exactly, as it depends on many factors such as the quality of the original recording, the type of sound being echoed, and the listener. However, in general one can expect the value of the offset Δt to be around one millisecond.

Since this scheme can only embed one bit in a signal, a practical approach consists of dividing the audio file into various blocks prior to the encoding process. Then each block is used to encode a bit, with the method described above. Moreover, if consecutive blocks are separated by a random number of unused samples, the detection and removal of the watermark becomes more difficult (Johnson & Katzenbeisser, 2000). Finally, all the blocks are concatenated back, and the watermarked audio file A' is created. This technique results in an embedding rate of around 16 bits per second without any degradation of the signal. Moreover, in some cases the resonance can even create a richer sound.

For watermark recovery, a technique known as *cepstrum autocorrelation* is used (Czerwinski et al., 1999). This technique produces a signal with two pronounced amplitude humps or *spikes*. By measuring the distance between these two spikes, one can determine if a one or a zero was initially encoded in the signal. This recovery process has the benefit that the original audio file A is not needed. However, this benefit also becomes a drawback in that the scheme presented here is susceptible to attack. This will be further explained in the sixth section.

Phase Coding

It is known that the human auditory system is less sensitive to the phase components of sound than to the noise components, a property that is exploited by some audio compression schemes. Phase coding (or phase distortion) makes use of this characteristic as well (Bender et al., 1996; Johnson & Katzenbeisser, 2000).

The method works by substituting the phase of the original audio signal A with one of two reference phases, each one encoding a bit of information. That is, the watermark data W is represented by a phase shift in the phase of A.

The original signal A is split into a series of short sequences A_i, each one of length l. Then a discrete Fourier transform (DFT) is applied to each one of the resulting segments. This transforms the signal representation from the time domain to the frequency domain, thus generating a matrix of phases Φ and a matrix of Fourier transform magnitudes.

The phase shifts between consecutive signal segments must be preserved in the watermarked

file *A′*. This is necessary because the human auditory system is very sensitive to relative phase differences, but not to absolute phase changes. In other words, the phase coding method works by substituting the phase of the initial audio segment with a reference phase that represents the data. After this, the phase of subsequent segments is adjusted in order to preserve the relative phases between them (Bender et al., 1996).

Given this, the embedding process inserts the watermark information in the phase vector of the first segment of *A*, namely $\vec{\Phi}_0$. Then it creates a new phase matrix Φ′, using the original phase differences found in Φ.

After this step, the original matrix of Fourier transform magnitudes is used alongside the new phase matrix Φ′ to construct the watermarked audio signal *A′*, by applying the inverse Fourier transform (that is, converting the signal back to the time domain). At this point, the absolute phases of the signal have been modified, but their relative differences are preserved. Throughout the process, the matrix of Fourier amplitudes remains constant. Any modifications to it could generate intolerable degradation (Dugelay & Roche, 2000).

In order to recover the watermark, the length of the segments, the DFT points, and the data interval must be known at the receiver. When the signal is divided into the same segments that were used for the embedding process, the following step is to calculate the DFT for each one of these segments. Once the transformation has been applied, the recovery process can measure the value of vector $\vec{\Phi}_0$ and thereby restore the originally encoded value for *W*.

With phase coding, an embedding rate between eight and 32 bits per second is possible, depending on the audio context. The higher rates are usually achieved when there is a noisy background in the audio signal. A higher embedding rate can result in phase dispersion, a distortion[13] caused by a break in the relationship of the phases between each of the frequency components (Bender et al., 1996).

Spread Spectrum Watermarking

Spread spectrum techniques for watermarking borrow most of the theory from the communications community (Czerwinski et al., 1999). The main idea is to embed a narrow-band signal (the watermark) into a wide-band channel (the audio file). The characteristics of both *A* and *W* seems to suit this model perfectly. In addition, spread spectrum techniques offer the possibility of protecting the watermark privacy by using a secret key to control the pseudorandom sequence generator that is needed in the process.

Generally, the message used as the watermark is a narrow band signal compared to the wide band of the cover (Dugelay & Roche, 2000; Kirovski & Malvar, 2001). Spread spectrum techniques allow the frequency bands to be matched before embedding the message. Furthermore, high frequencies are relevant for the invisibility of the watermark but are inefficient as far as robustness is concerned, whereas low frequencies have the opposite characteristics. If a low energy signal is embedded on each of the frequency bands, this conflict is partially solved. This is why spread spectrum techniques are valuable not only for robust communication but for watermarking as well.

There are two basic approaches to spread spectrum techniques: *direct sequence* and *frequency hopping*. In both of these approaches the idea is to spread the watermark data across a large frequency band, namely the entire audible spectrum.

In the case of direct sequence, the cover signal *A* is modulated by the watermark message *m* and a pseudorandom (PN) noise sequence, which has a wide frequency spectrum. As a consequence, the spectrum of the resulting message *m′* is spread over the available band. Then, the spread message *m′* is attenuated in order to obtain the watermark *W*. This watermark is then added to the original file, for example as additive random noise, in order to obtain the watermarked version *A′*. To keep the noise level down, the attenuation

performed to *m′* should yield a signal with about 0.5% of the dynamic range of the cover file *A* (Bender et al., 1996).

In order to recover the watermark, the watermarked audio signal *A′* is modulated with the PN-sequence to remove it. The demodulated signal is then *W*. However, some keying mechanisms can be used when embedding the watermark, which means that at the recovery end a detector must also be used. For example, if bi-phase shift keying is used when embedding *W*, then a phase detector must be used at the recovery process (Czerwinski et al., 1999).

In the case of *frequency hopping*, the cover frequency is altered using a random process, thus describing a wide range of frequency values. That is, the frequency-hopping method selects a pseudorandom subset of the data to be watermarked. The watermark *W* is then attenuated and merged with the selected data using one of the methods explained in this chapter, such as coefficient quantization in a transform domain. As a result, the modulated watermark has a wide spectrum.

For the detection process, the pseudorandom generator used to alter the cover frequency is used to recover the parts of the signal where the watermark is hidden. Then the watermark can be recovered by using the detection method that corresponds to the embedding mechanism used.

A crucial factor for the performance of spread spectrum techniques is the synchronization between the watermarked audio signal *A′* and the PN-sequence (Dugelay & Roche, 2000; Kirovski & Malvar, 2001). This is why the particular PN-sequence used acts as a key to the recovery process. Nonetheless, some attacks can focus on this delicate aspect of the model.

MEASURING FIDELITY

Artists, and digital content owners in general, have many reasons for embedding watermarks in their copyrighted works. These reasons have been stated in the previous sections. However, there is a big risk in performing such an operation, as the quality of the musical content might be degraded to a point where its value is diminished. Fortunately, the opposite is also possible and, if done right, digital watermarks can add value to content (Acken, 1998).

Content owners are generally concerned with the degradation of the cover signal quality, even more than users of the content (Craver, Yeo, & Yeung, 1998). They have access to the unwatermarked content with which to compare their audio files. Moreover, they have to decide between the amount of tolerance in quality degradation from the watermarking process and the level of protection that is achieved by embedding a stronger signal. As a restriction, an embedded watermark has to be detectable in order to be valuable.

Given this situation, it becomes necessary to measure the impact that a marking scheme has on an audio signal. This is done by measuring the fidelity of the watermarked audio signal *A′*, and constitutes the first measure that is defined in this chapter.

As fidelity refers to the similitude between an original and a watermarked signal, a statistical metric must be used. Such a metric will fall in one of two categories: difference metrics or correlation metrics.

Difference metrics, as the name states, measure the difference between the undistorted original audio signal *A* and the distorted watermarked signal *A′*. The popularity of these metrics is derived from their simplicity (Kutter & Petitcolas, 1999). In the case of digital audio, the most common difference metric used for quality evaluation of watermarks is the *signal to noise ratio* (SNR). This is usually measured in decibels (dB), so SNR(dB) = 10 $\log_{10}$ (SNR).

The signal to noise ratio, measured in decibels, is defined by the formula:

$$SNR(dB) = 10\log_{10}\frac{\sum_n A_n^2}{\sum_n (A_n - A'_n)^2}$$

where A_n corresponds to the n^{th} sample of the original audio file A, and A'_n to the n^{th} sample of the watermarked signal A'. This is a measure of quality that reflects the quantity of distortion that a watermark imposes on a signal (Gordy & Burton, 2000).

Another common difference metric is the *peak signal to noise ratio* (PSNR), which measures the maximum signal to noise ratio found on an audio signal. The formula for the PSNR, along with some other difference metrics found in the literature are presented in Table 1 (Kutter & Hartung, 2000; Kutter & Petitcolas, 1999).

Although the tolerable amount of noise depends on both the watermarking application and the characteristics of the unwatermarked audio signal, one could expect to have perceptible noise distortion for SNR values of 35dB (Petitcolas & Anderson, 1999).

Correlation metrics measure distortion based on the statistical correlation between the original and modified signals. They are not as popular as the difference distortion metrics, but it is important to state their existence. Table 2 shows the most important of these.

For the purpose of audio watermark benchmarking, the use of the *signal to noise ratio* (SNR) should be used to measure the fidelity of the watermarked signal with respect to the original. This decision follows most of the literature that deals with the topic (Gordy & Burton, 2000; Kutter & Petitcolas, 1999, 2000; Petitcolas & Anderson,

Table 1. Common difference distortion metrics

Maximum Difference	$MD = \max \mid A_n - A'_n \mid$
Average Absolute Difference	$AD = \frac{1}{N}\sum_n \mid A_n - A'_n \mid$
Normalized Average Absolute Difference	$NAD = \sum_n \mid A_n - A'_n \mid / \sum_n \mid A_n \mid$
Mean Square Error	$MSE = \frac{1}{N}\sum_n (A_n - A'_n)^2$
Normalized Mean Square Error	$NMSE = \sum_n (A_n - A'_n)^2 / \sum_n A_n^2$
LP-Norm	$LP = \left(\frac{1}{N}\sum_n \mid A_n - A'_n \mid \right)^{1/p}$
Laplacian Mean Square Error	$LMSE = \sum_n (\nabla^2 A_n - \nabla^2 A'_n)^2 / \sum_n (\nabla^2 A_n)^2$
Signal to Noise Ratio	$SNR = \sum_n A_n^2 / \sum_n (A_n - A'_n)^2$
Peak Signal to Noise Ratio	$PSNR = N \max_n A_n^2 / \sum_n (A_n - A'_n)^2$
Audio Fidelity	$AF = 1 - \sum_n (A_n - A'_n)^2 / \sum_n A_n^2$

Table 2. Correlation distortion metrics

Normalized Cross-Correlation	$NC = \sum_n A_n \tilde{A}_n / \sum_n A_n^2$
Correlation Quality	$CQ = \sum_n A_n \tilde{A}_n / \sum_n A_n$

1999). Nonetheless, in this measure the term *noise* refers to statistical noise, or a deviation from the original signal, rather than to perceived noise on the side of the hearer. This result is due to the fact that the SNR is not well correlated with the human auditory system (Kutter & Hartung, 2000). Given this characteristic, the effect of perceptual noise needs to be addressed later.

In addition, when a metric that outputs results in decibels is used, comparisons are difficult to make, as the scale is not linear but rather logarithmic. This means that it is more useful to present the results using a normalized quality rating. The ITU-R Rec. 500 quality rating is perfectly suited for this task, as it gives a quality rating on a scale of 1 to 5 (Arnold, 2000; Piron et al., 1999). Table 3 shows the rating scale, along with the quality level being represented.

This quality rating is computed by using the formula:

$$Quality = F = \frac{5}{1 + N * SNR}$$

where N is a normalization constant and SNR is the measured signal to noise ratio. The resulting value corresponds to the fidelity F of the watermarked signal.

Table 3. ITU-R Rec. 500 quality rating

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Data Payload

The fidelity of a watermarked signal depends on the amount of embedded information, the strength of the mark, and the characteristics of the host signal. This means that a comparison between different algorithms must be made under equal conditions. That is, while keeping the payload fixed, the fidelity must be measured on the same audio cover signal for all watermarking techniques being evaluated.

However, the process just described constitutes a single measure event and will not be representative of the characteristics of the algorithms being evaluated, as results can be biased depending on the chosen parameters. For this reason, it is important to perform the tests using a variety of audio signals, with changing size and nature (Kutter & Petitcolas, 2000). Moreover, the test should also be repeated using different keys.

The amount of information that should be embedded is not easy to determine, and depends on the application of the watermarking scheme. In Kutter and Petitcolas (2000) a message length of 100 bits is used on their test of image watermarking systems as a representative value. However, some secure watermarking protocols might need a bigger payload value, as the watermark W could include a cryptographic signature for both the audio file A, and the watermark message m in order to be more secure (Katzenbeisser & Veith, 2002). Given this, it is recommended to use a longer watermark bitstream for the test, so that a

real world scenario is represented. A watermark size of 128 bits is big enough to include two 56-bit signatures and a unique identification number that identifies the owner.

Speed

Besides fidelity, the content owner might be interested in the time it takes for an algorithm to embed a mark (Gordy & Burton, 2000). Although speed is dependent on the type of implementation (hardware or software), one can suppose that the evaluation will be performed on software versions of the algorithms. In this case, it is a good practice to perform the test on a machine with similar characteristics to the one used by the end user (Petitcolas, 2000). Depending on the application, the value for the time it takes to embed a watermark will be incorporated into the results of the test. This will be done later, when all the measures are combined together.

MEASURING ROBUSTNESS

Watermarks have to be able to withstand a series of signal operations that are performed either intentionally or unintentionally on the cover signal and that can affect the recovery process. Given this, watermark designers try to guarantee a minimum level of robustness against such operations. Nonetheless, the concept of robustness is ambiguous most of the time and thus claims about a watermarking scheme being robust are difficult to prove due to the lack of testing standards (Craver, Perrig, & Petitcolas, 2000).

By defining a standard metric for watermark robustness, one can then assure fairness when comparing different technologies. It becomes necessary to create a detailed and thorough test for measuring the ability that a watermark has to withstand a set of clearly defined signal operations. In this section these signal operations are presented, and a practical measure for robustness is proposed.

How to Measure

Before defining a metric, it must be stated that one does not need to erase a watermark in order to render it useless. It is said that a watermarking scheme is robust when it is able to withstand a series of attacks that try to degrade the quality of the embedded watermark, up to the point where it is removed, or its recovery process is unsuccessful. This means that just by interfering with the detection process a person can create a successful attack over the system, even unintentionally.

However, in some cases one can overcome this characteristic by using error-correcting codes or a stronger detector (Cox et al., 2002). If an error correction code is applied to the watermark message, then it is unnecessary to entirely recover the watermark W in order to successfully retrieve the embedded message m. The use of stronger detectors can also be very helpful in these situations. For example, if a marking scheme has a publicly available detector, then an attacker will try to tamper with the cover signal up to the point where the detector does not recognize the watermark's presence[14]. Nonetheless, the content owner may have another version of the watermark detector, one that can successfully recover the mark after some extra set of signal processing operations. This "special" detector might not be released for public use for economic, efficiency or security reasons. For example, it might only be used in court cases. The only thing that is really important is that it is possible to design a system with different detector strengths.

Given these two facts, it makes sense to use a metric that allows for different levels of robustness, instead of one that only allows for two different states (the watermark is either robust or not). With this characteristic in mind, the basic procedure for measuring robustness is a three-step process, defined as follows:

1. For each audio file in a determined test set embed a random watermark **W** on the

audio signal A, with the maximum strength possible that does not diminish the fidelity of the cover below a specified minimum (Petitcolas & Anderson, 1999).

2. Apply a set of relevant signal processing operations to the watermarked audio signal **A'.**
3. Finally, for each audio cover, extract the watermark **W** using the corresponding detector and measure the success of the recovery process.

Some of the early literature considered the recovery process successful only if the whole watermark message *m* was recovered (Petitcolas, 2000; Petitcolas & Anderson, 1999). This was in fact a binary robustness metric. However, the use of the *bit-error rate* has become common recently (Gordy & Burton, 2000; Kutter & Hartung, 2000; Kutter & Petitcolas, 2000), as it allows for a more detailed scale of values. The *bit-error rate* (BER) is defined as the ratio of incorrect extracted bits to the total number of embedded bits and can be expressed using the formula:

$$BER = \frac{100}{l}\sum_{n=0}^{l-1}\begin{cases}1, & W'_n = W_n \\ 0, & W'_n \neq W_n\end{cases}$$

where l is the watermark length, W_n corresponds to the n^{th} bit of the embedded watermark and W'_n corresponds to the n^{th} bit of the recovered watermark. In other words, this measure of robustness is the certainty of detection of the embedded mark (Arnold, 2000). It is easy to see why this measure makes more sense, and thus should be used as the metric when evaluating the success of the watermark recovery process and therefore the robustness of an audio watermarking scheme.

A final recommendation must be made at this point. The three-step procedure just described should be repeated several times, since the embedded watermark *W* is randomly generated and the recovery can be successful by chance (Petitcolas, 2000).

Up to this point no details have been given about the signal operations that should be performed in the second step of the robustness test. As a rule of thumb, one should include as a minimum the operations that the audio cover is expected to go through in a real world application. However, this will not provide enough testing, as a malicious attacker will most likely have access to a wide range of tools as well as a broad range of skills. Given this situation, several scenarios should be covered. In the following sections the most common signal operations and attacks that an audio watermark should be able to withstand are presented.

Audio Restoration Attack

Audio restoration techniques have been used for several years now, specifically for restoring old audio recordings that have audible artifacts. In audio restoration the recording is digitized and then analyzed for degradations. After these degradations have been localized, the corresponding samples are eliminated. Finally the recording is reconstructed (that is, the missing samples are recreated) by interpolating the signal using the remaining samples.

One can assume that the audio signal is the product of a stationary autoregressive (AR) process of finite order (Petitcolas & Anderson, 1998). With this assumption in mind, one can use an audio segment to estimate a set of AR parameters and then calculate an approximate value for the missing samples. Both of the estimates are calculated using a least-square minimization technique.

Using the audio restoration method just described one can try to render a watermark undetectable by processing the marked audio signal *A'*. The process is as follows: First divide the audio signal *A'* into *N* blocks of size *m* samples each. A value of *m*=1000 samples has been proposed in the literature (Petitcolas & Anderson, 1999). A block of length *l* is removed from the middle of each block and then restored using the AR audio

restoration algorithm. This generates a reconstructed block also of size *m*. After the *N* blocks have been processed they are concatenated again, and an audio signal *B′* is produced. It is expected that *B′* will be closer to *A* than to *A′* and thus the watermark detector will not find any mark in it.

An error free restoration is theoretically possible in some cases, but this is not desired since it would produce a signal identical to *A′*. What is expected is to create a signal that has an error value big enough to mislead the watermark detector, but small enough to prevent the introduction of audible noise. Adjusting the value of the parameter *l* controls the magnitude of the error (Petitcolas & Anderson, 1999). In particular, a value of *l*=80 samples has proven to give good results.

Invertibility Attack

When resolving ownership cases in court, the disputing parties can both claim that they have inserted a valid watermark on the audio file, as it is sometimes possible to embed multiple marks on a single cover signal. Clearly, one mark must have been embedded before the other.

The ownership is resolved when the parties are asked to show the original work to court. If Alice has the original audio file *A*, which has been kept stored in a safe place, and Mallory has a counterfeit original file $\tilde{A}$, which has been derived from *A*, then Alice can search for her watermark *W* in Mallory's file and will most likely find it. The converse will not happen, and the case will be resolved (Craver et al., 2000). However, an attack to this procedure can be created, and is known as an *invertibility attack*.

Normally the content owner adds a watermark *W* to the audio file *A*, creating a watermarked audio file $A' = A+W$, where the sign "+" denotes the embedding operation. This file is released to the public, while the original *A* and the watermark *W* are stored in a safe place. When a suspicious audio file $\tilde{A}$ appears, the difference $\tilde{W} = \tilde{A} - A$ is computed. This difference should be equal to *W* if *A′* and $\tilde{A}$ are equal, and very close to *W* if $\tilde{A}$ was derived from *A′*. In general, a correlation function $f(W, \tilde{W})$ is used to determine the similarity between the watermark *W* and the extracted data $\tilde{W}$. This function will yield a value close to 1, if *W* and $\tilde{W}$ are similar.

However, Mallory can do the following: she can subtract (rather than add) a second watermark *w* from Alice's watermarked file *A′*, using the inverse of the embedding algorithm. This yields an audio file $\hat{A} = A' - w = A + W - w$, which Mallory can now claim to be the original audio file, along with *w* as the original watermark (Craver, Memon, Yeo, & Yeung, 1998). Now both Alice and Mallory can claim copyright violation from their counterparts.

When the two originals are compared in court, Alice will find that her watermark is present in Mallory's audio file, since $\hat{A} - A = W - w$ is calculated, and $f(W-w, W) \approx 1$. However, Mallory can show that when $A - \hat{A} = w - W$ is calculated, then $f(w-W, w) \approx 1$ as well. In other words, Mallory can show that her mark is also present in Alice's work, even though Alice has kept it locked at all times (Craver, Memon, & Yeung, 1996; Craver, Yeo et al., 1998). Given the symmetry of the equations, it is impossible to decide who is the real owner of the original file. A deadlock is thus created (Craver, Yeo et al., 1998; Pereira et al., 2001).

This attack is a clear example of how one can render a mark unusable without having to remove it, by exploiting the invertibility of the watermarking method, which allows an attacker to remove as well as add watermarks. Such an attack can be prevented by using a non-invertible cryptographic signature in the watermark *W;* that is, using a secure watermarking protocol (Katzenbeisser & Veith, 2002; Voloshynovskiy, Pereira, Pun et al., 2001).

Specific Attack on Echo Watermarking

The echo watermarking technique presented in this chapter can be easily "attacked" simply by detecting the echo and then removing the delayed signal by inverting the convolution formula that was used to embed it. However, the problem consists of detecting the echo without knowing the original signal and the possible delay values. This problem is referred to as *blind echo cancellation*, and is known to be difficult to solve (Petitcolas, Anderson, & G., 1998). Nonetheless, a practical solution to this problem appears to lie in the same function that is used for echo watermarking extraction: *cepstrum autocorrelation*. Cepstrum analysis, along with a brute force search can be used together to find the echo signal in the watermarked audio file A'.

A detailed description of the attack is given by Craver et al. (2000), and the idea is as follows: If we take the power spectrum of $A'(t) = A(t) + \alpha A(t - \Delta t)$, denoted by Φ and then calculate the logarithm of Φ, the amplitude of the delayed signal can be augmented using an autocovariance function[15] over the power spectrum $\Phi'(\ln(\Phi))$. Once the amplitude has been increased, then the "hump" of the signal becomes more visible and the value of the delay Δt can be determined (Petitcolas et al., 1998).

Experiments show that when an artificial echo is added to the signal, this attack works well for values of Δt between 0.5 and three milliseconds (Craver et al., 2000). Given that the watermark is usually embedded with a delay value that ranges from 0.5 to two milliseconds, this attack seems to be well suited for the technique and thus very likely to be successful (Petitcolas et al., 1999).

Collusion Attack

A collusion attack, also known as *averaging*, is especially effective against basic fingerprinting schemes. The basic idea is to take a large number of watermarked copies of the same audio file, and average them in order to produce an audio signal without a detectable mark (Craver et al., 2000; Kirovski & Malvar, 2001).

Another possible scenario is to have copies of multiple works that have been embedded with the same watermark. By averaging the sample values of the audio signals, one could estimate the value of the embedded mark, and then try to subtract it from any of the watermarked works. It has been shown that a small number (around 10) of different copies are needed in order to perform a successful collusion attack (Voloshynovskiy, Pereira, Pun et al., 2001). An obvious countermeasure to this attack is to embed more than one mark on each audio cover, and to make the marks dependant on the characteristics of the audio file itself (Craver et al., 2000).

Signal Diminishment Attacks and Common Processing Operations

Watermarks must be able to survive a series of signal processing operations that are commonly performed on the audio cover work, either intentionally or unintentionally. Any manipulation of an audio signal can result in a successful removal of the embedded mark. Furthermore, the availability of advanced audio editing tools on the Internet, such as Audacity (Dannenberg & Mazzoni, 2002), implies that these operations can be performed without an extensive knowledge of digital signal processing techniques. The removal of a watermark by performing one of these operations is known as a signal diminishment attack, and probably constitutes the most common attack performed on digital watermarks (Meerwald & Pereira, 2002).

Given this, a set of the most common signal operations must be specified, and watermark resistance to these must be evaluated. Even though an audio file will most likely not be subject to all the possible operations, a thorough list is necessary. Defining which subset of these operations

is relevant for a particular watermarking scheme is a task that needs to be done; however, this will be addressed later in the chapter.

The signal processing operations presented here are classified into eight different groups, according to the presentation made in Petitcolas et al. (2001). These are:

- **Dynamics.** These operations change the loudness profile of the audio signal. The most basic way of performing this consists of increasing or decreasing the loudness directly. More complicated operations include limiting, expansion and compression, as they constitute nonlinear operations that are dependant on the audio cover.
- **Filter.** Filters cut off or increase a selected part of the audio spectrum. Equalizers can be seen as filters, as they increase some parts of the spectrum, while decreasing others. More specialized filters include low-pass, high-pass, all-pass, FIR, and so forth.
- **Ambience.** These operations try to simulate the effect of listening to an audio signal in a room. Reverb and delay filters are used for this purpose, as they can be adjusted in order to simulate the different sizes and characteristics that a room can have.
- **Conversion.** Digital audio files are nowadays subject to format changes. For example, old monophonic signals might be converted to stereo format for broadcast transmission. Changes from digital to analog representation and back are also common, and might induce significant quantization noise, as no conversion is perfect.
- **Lossy compression** algorithms are becoming popular, as they reduce the amount of data needed to represent an audio signal. This means that less bandwidth is needed to transmit the signal, and that less space is needed for its storage. These compression algorithms are based on psychoacoustic models and, although different implementations exist, most of them rely on deleting information that is not perceived by the listener. This can pose a serious problem to some watermarking schemes, as they sometimes will hide the watermark exactly in these imperceptible regions. If the watermarking algorithm selects these regions using the same method as the compression algorithm, then one just needs to apply the lossy compression algorithm to the watermarked signal in order to remove the watermark.
- **Noise** can be added in order to remove a watermark. This noise can even be imperceptible, if it is shaped to match the properties of the cover signal. Fragile watermarks are especially vulnerable to this attack. Sometimes noise will appear as the product of other signal operations, rather than intentionally.
- **Modulation** effects like vibrato, chorus, amplitude modulation and flanging are not common post-production operations. However, they are included in most of the audio editing software packages and thus can be easily used in order to remove a watermark.
- **Time stretch** and pitch shift. These operations either change the length of an audio passage without changing its pitch, or change the pitch without changing its length in time. The use of time stretch techniques has become common in radio broadcasts, where stations have been able to increase the number of advertisements without devoting more air time to these (Kuczynski, 2000).
- **Sample permutations.** This group consists of specialized algorithms for audio manipulation, such as the attack on echo hiding just presented. Dropping of some samples in order to misalign the watermark decoder is also a common attack to spread-spectrum watermarking techniques.

It is not always clear how much processing a watermark should be able to withstand. That is, the specific parameters of the diverse filtering operations that can be performed on the cover signal are not easy to determine. In general terms one could expect a marking scheme to be able to survive several processing operations up to the point where they introduce annoying audible effects on the audio work. However, this rule of thumb is still too vague.

Fortunately, guidelines and minimum requirements for audio watermarking schemes have been proposed by different organizations such as the Secure Digital Music Initiative (SDMI), International Federation of the Phonographic Industry (IFPI), and the Japanese Society for Rights of Authors, Composers and Publishers (JASRAC). These guidelines constitute the baseline for any robustness test. In other words, they describe the minimum processing that an audio watermark should be able to resist, regardless of its intended application. Table 4 summarizes these requirements (JASRAC, 2001; SDMI, 2000).

False Positives

When testing for false positives, two different scenarios must be evaluated. The first one occurs when the watermark detector signals the presence of a mark on an unmarked audio file. The second case corresponds to the detector successfully finding a watermark W' on an audio file that has been marked with a watermark W (Cox et al., 2002; Kutter & Hartung, 2000; Petitcolas et al., 2001).

Table 4. Summary of SDMI, STEP and IFPI requirements

<table>
<tr><th>Processing Operation</th><th>Requirements</th></tr>
<tr><td>Digital to Analog conversion</td><td>Two consecutive digital to analog and analog to digital conversions.</td></tr>
<tr><td>Equalization</td><td>10 band graphic equalizer with the following characteristics:
<table>
<tr><td>Freq. (Hz)</td><td>31</td><td>62</td><td>125</td><td>250</td><td>500</td><td>1k</td><td>2k</td><td>4k</td><td>8k</td><td>16k</td></tr>
<tr><td>Gain (db)</td><td>-6</td><td>+6</td><td>-6</td><td>+3</td><td>-6</td><td>+6</td><td>-6</td><td>+6</td><td>-6</td><td>+6</td></tr>
</table></td></tr>
<tr><td>Band-pass filtering</td><td>100 Hz – 6 kHz, 12dB/oct.</td></tr>
<tr><td>Time stretch and pitch change</td><td>+/- 10% compression and decompression</td></tr>
<tr><td>Codecs (at typically used data rates)</td><td>AAC, MPEG-4 AAC with perceptual noise substitution, MPEG-1 Audio Layer 3, Q-Design, Windows Media Audio, Twin-VQ, ATRAC-3, Dolby Digital AC-3, ePAC, RealAudio, FM, AM, PCM</td></tr>
<tr><td>Noise addition</td><td>Adding white noise with constant level of 40dB lower than total averaged music power (SNR: 40dB)</td></tr>
<tr><td>Time scale modification</td><td>Pitch invariant time scaling of +/- 4%</td></tr>
<tr><td>Wow and flutter</td><td>0.5% rms, from DC to 250Hz</td></tr>
<tr><td>Echo addition</td><td>Delay up to 100 milliseconds, feedback coefficient up to 0.5</td></tr>
<tr><td>Down mixing and surround sound processing</td><td>Stereo to mono, 6 channel to stereo, SRS, spatializer, Dolby surround, Dolby headphone</td></tr>
<tr><td>Sample rate conversion</td><td>44.1 kHz to 16 kHz, 48 kHz to 44.1 kHz, 96 kHz to 48/44.1 kHz</td></tr>
<tr><td>Dynamic range reduction</td><td>Threshold of 50dB, 16dB maximum compression
Rate: 10 millisecond attack, 3 second recovery</td></tr>
<tr><td>Amplitude compression</td><td>16 bits to 8 bits</td></tr>
</table>

The testing procedure for both types of false positives is simple. In the first case one just needs to run the detector on a set of unwatermarked works. For the second case, one can embed a watermark W using a given key K, and then try to extract a different mark W' while using the same key K. The *false positive rate* (FPR) is then defined as the number of successful test runs divided by the total number of test runs. A successful test run is said to occur whenever a false positive is detected.

However, a big problem arises when one takes into account the required false positive rate for some schemes. For example, a popular application such as DVD watermarking requires a false positive rate of 1 in 10^{12} (Cox et al., 2002). In order to verify that this rate is accomplished one would need to run the described experiment during several years. Other applications such as proof of ownership in court are rare, and thus require a lower false positive rate. Nonetheless, a false rate probability of 10^{-6}, required for the mentioned application, can be difficult to test.

MEASURING PERCEPTIBILITY

Digital content consumers are aware of many aspects of emerging watermarking technologies. However, only one prevails over all of them: users are concerned with the appearance of perceptible (audible) artifacts due to the use of a watermarking scheme. Watermarks are supposed to be imperceptible (Cox et al., 2002). Given this fact, one must carefully measure the amount of distortion that the listener will perceive on a watermarked audio file, as compared to its unmarked counterpart. Formal listening tests have been considered the only relevant method for judging audio quality, as traditional objective measures such as the signal-to-noise ratio (SNR) or total-harmonic-distortion[16] (THD) have never been shown to reliably relate to the perceived audio quality, as they can not be used to distinguish inaudible artifacts from audible noise (ITU, 2001; Kutter & Hartung, 2000; Thiede & Kabot, 1996). There is a need to adopt an objective measurement test for perceptibility of audio watermarking schemes.

Furthermore, one must be careful, as perceptibility must not be viewed as a binary condition (Arnold & Schilz, 2002; Cox et al., 2002). Different levels of perceptibility can be achieved by a watermarking scheme; that is, listeners will perceive the presence of the watermark in different ways. Auditory sensitivities vary significantly from individual to individual. As a consequence, any measure of perceptibility that is not binary should accurately reflect the probability of the watermark being detected by a listener.

In this section a practical and automated evaluation of watermark perceptibility is proposed. In order to do so, the human auditory system (HAS) is first described. Then a formal listening test is presented, and finally a psychoacoustical model for automation of such a procedure is outlined.

Human Auditory System (HAS)

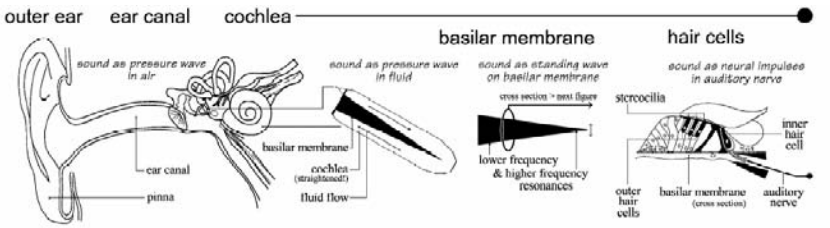
Figure 4, taken from Robinson (2002), presents the physiology of the human auditory system. Each one of its components is now described.

The *pinna* directionally filters incoming sounds, producing a spectral coloration known as *head related transfer function* (or HRTF). This function enables human listeners to localize the sound source in three dimensions.

The *ear canal* filters the sound, attenuating both low and high frequencies. As a result, a resonance arises around 5 kHz. After this, small bones known as the *timpanic membrane* (or ear drum), *malleus* and *incus* transmit the sound pressure wave through the middle ear. The outer and middle ear perform a band pass filter operation on the input signal.

The sound wave arrives at the fluid-filled *cochlea*, a coil within the ear that is partially protected by a bone. Inside the cochlea resides the *basilar membrane* (BM), which semi-divides

Figure 4. Overview of the human auditory system (HAS)



it. The basilar membrane acts as a spectrum analyzer, as it divides the signal into frequency components. Each point on the membrane resonates at a different frequency, and the spacing of these resonant frequencies along the BM is almost logarithmic. The effective frequency selectivity is related to the width of the filter characteristic at each point.

The *outer hair cells*, distributed along the length of the BM, react to feedback from the brainstem. They alter their length to change the resonant properties of the BM. As a consequence, the frequency response of the membrane becomes amplitude dependent.

Finally, the inner hair cells of the basilar membrane fire when the BM moves upward. In doing so, they transduce the sound wave at each point into a signal on the auditory nerve. In this way the signal is half wave rectified. Each cell needs a certain time to recover between successive firings, so the average response during a steady tone is lower than at its onset. This means that the inner hair cells act as an automatic gain control.

The net result of the process described above is that an audio signal, which has a relatively wide-bandwidth, and large dynamic range, is encoded for transmission along the nerves. Each one of these nerves offers a much narrower bandwidth, and limited dynamic range. In addition, a critical process has happened during these steps. Any information that is lost due to the transduction process within the cochlea is not available to the brain. In other words, the cochlea acts as a lossy coder. The vast majority of what we cannot hear is attributable to this transduction process (Robinson & Hawksford, 1999).

Detailed modeling of the components and processes just described will be necessary when creating an auditory model for the evaluation of watermarked audio. In fact, by representing the audio signal at the basilar membrane, one can effectively model what is effectively perceived by a human listener.

Perceptual Phenomena

As was just stated, one can model the processes that take place inside the HAS in order to represent how a listener responds to auditory stimuli. Given its characteristics, the HAS responds differently depending on the frequency and loudness of the input. This means that all components of a watermark may not be equally perceptible. Moreover, it also denotes the need of using a perceptual model to effectively measure the amount of distortion that is imposed on an audio signal when a mark is embedded. Given this fact, in this section the main processes that need to be included on a perceptual model are presented.

Sensitivity refers to the ear's response to direct stimuli. In experiments designed to measure sensitivity, listeners are presented with isolated stimuli and their perception of these stimuli is tested. For example, a common test consists of measuring the minimum sound intensity required to hear a particular frequency (Cox et al., 2002). The main characteristics measured for sensitivity are *frequency* and *loudness.*

The responses of the HAS are frequency dependent; variations in frequency are perceived as different tones. Tests show that the ear is most sensitive to frequencies around 3kHz and that sensitivity declines at very low (20 Hz) and very high (20 kHz) frequencies.

Regarding loudness, different tests have been performed to measure sensitivity. As a general result, one can state that the HAS is able to discern smaller changes when the average intensity is louder. In other words, the human ear is more sensitive to changes in louder signals than in quieter ones.

The second phenomenon that needs to be taken into account is *masking.* A signal that is clearly audible if presented alone can be completely inaudible in the presence of another signal, the masker. This effect is known as masking, and the masked signal is called the *maskee*. For example, a tone might become inaudible in the presence of a second tone at a nearby frequency that is louder. In other words, masking is a measure of a listener's response to one stimulus in the presence of another.

Two different kinds of masking can occur: simultaneous masking and temporal masking (Swanson et al., 1998). In simultaneous masking, both the masker and the maskee are presented at the same time and are quasi-stationary (ITU, 2001). If the masker has a discrete bandwidth, the threshold of hearing is raised even for frequencies below or above the masker. In the situation where a noise-like signal is masking a tonal signal, the amount of masking is almost frequency independent; if the sound pressure of the maskee is about 5 dB below that of the masker, then it becomes inaudible. For other cases, the amount of masking depends on the frequency of the masker.

In temporal masking, the masker and the maskee are presented at different times. Shortly after the decay of a masker, the masked threshold is closer to simultaneous masking of this masker than to the absolute threshold (ITU, 2001). Depending on the duration of the masker, the decay time of the threshold can vary between five ms and 150 ms. Furthermore, weak signals just before loud signals are masked. The duration of this backward masking effect is about five ms.

The third effect that has to be considered is *pooling.* When multiple frequencies are changed rather than just one, it is necessary to know how to combine the sensitivity and masking information for each frequency. Combining the perceptibilities of separate distortions gives a single estimate for the overall change in the work. This is known as pooling. In order to calculate this phenomenon, it is common to apply the formula:

$$D(A, A') = \left(\sum_i | d[i] |^p \right)^{\frac{1}{p}}$$

where d[i] is an estimate of the likelihood that an individual will notice the difference between A and A' in a temporal sample (Cox et al., 2002). In the case of audio, a value of p=1 is sometimes appropriate, which turns the equation into a linear summation.

ABX Listening Test

Audio quality is usually evaluated by performing a listening test. In particular, the ABX listening test is commonly used when evaluating the quality of watermarked signals. Other tests for audio watermark quality evaluation, such as the one described in Arnold and Schilz (2002), follow a similar methodology as well. Given this, it becomes desirable to create an automatic model that predicts the response observed from a human listener in such a procedure.

In an ABX test the listener is presented with three different audio clips: selection A (in this case the non-watermarked audio), selection B (the watermarked audio) and X (either the watermarked or non-watermarked audio), drawn at random. The listener is then asked to decide if selection X is equal to A or B. The number of correct answers is the basis to decide if the watermarked audio is perceptually different than the original audio

and one will, therefore, declare the watermarking algorithm as "perceptible". In the other case, if the watermarked audio is perceptually equal to the original audio, the watermarking algorithm will be declared as *transparent*, or imperceptible. In the particular case of Arnold and Schilz (2002), the level of transparency is assumed to be determined by the noise-to-mask ratio (NMR).

The ABX test is fully described in ITU Recommendation ITU-R BS.1116, and has been successfully used for subjective measurement of impaired audio signals. Normally only one attribute is used for quality evaluation. It is also defined that this attribute represents any and all detected differences between the original signal and the signal under test. It is known as *basic audio quality* (BAQ), and is calculated as the difference between the grade given to the impaired signal and the grade given to the original signal. Each one of these grades uses the five-level impairment scale that was presented previously. Given this fact, values for the BAQ range between 0 and -4, where 0 corresponds to an imperceptible impairment and -4 to one judged as very annoying.

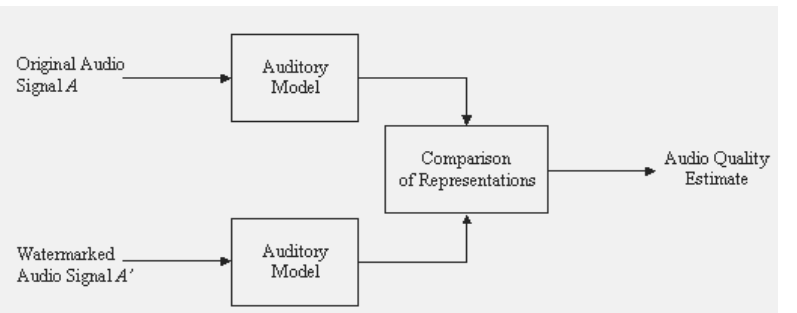
Although its results are highly reliable, there are many problems related to performing an ABX test for watermark quality evaluation. One of them is the subjective nature of the test, as the perception conditions of the listener may vary with time. Another problem arises from the high costs associated with the test. These costs include the setup of audio equipment[17], construction of a noise-free listening room, and the costs of employing individuals with extraordinarily acute hearing. Finally, the time required to perform extensive testing also poses a problem to this alternative.

Given these facts it becomes desirable to automate the ABX listening test, and incorporate it into a perceptual model of the HAS. If this is implemented, then the task measuring perceptibility can be fully automated and thus watermarking schemes can be effectively and thoroughly evaluated. Fortunately, several perceptual models for audio processing have been proposed. Specifically, in the field of audio coding, psychoacoustic models have been successfully implemented to evaluate the perceptual quality of coded audio. These models can be used as a baseline performance tool for measuring the perceptibility of audio watermarking schemes; thus they are now presented.

A Perceptual Model

A perceptual model used for evaluation of watermarked content must compare the quality of two different audio signals in a way that is similar to the ABX listening test. These two signals correspond to the original audio cover *A* and the watermarked audio file *A'*. An ideal system will receive both signals as an input, process them through an auditory model, and compare the representations given by this model (Thiede et al., 1998). Finally it will return a score for the watermarked file *A'*

Figure 5. Architecture of a perceptual measurement system



in the five-level impairment scale. More importantly, the results of such an objective test must be highly correlated with those achieved under a subjective listening test (ITU, 2001). The general architecture of such a perceptual measurement system is depicted in Figure 5.

The auditory model used to process the input signals will have a similar structure to that of the HAS. In general terms, the response of each one of the components of the HAS is modeled by a series of filters. In particular, a synopsis of the models proposed in Robinson and Hawksford (1999), Thiede and Kabot (1996), Thiede et al. (1998), and ITU (2001) is now presented.

The filtering performed by the pinna and ear canal is simulated by an FIR filter, which has been derived from experiments with a dummy head. More realistic approaches can use measurements from human subjects.

After this prefiltering, the audio signal has to be converted to a basilar membrane representation. That is, the amplitude dependent response of the basilar membrane needs to be simulated. In order to do this, the first step consists of processing the input signal through a bank of amplitude dependant filters, each one adapted to the frequency response of a point on the basilar membrane. The center frequency of each filter should be linearly spaced on the Bark scale, a commonly used frequency scale[18]. The actual number of filters to be used depends on the particular implementation. Other approaches might use a fast Fourier transform to decompose the signal, but this creates a trade-off between temporal and spectral resolution (Thiede & Kabot, 1996).

At each point in the basilar membrane, its movement is transduced into an electrical signal by the hair cells. The firing of individual cells is pseudorandom, but when the individual signals are combined, the proper motion of the BM is derived. Simulating the individual response of each hair cell and combining these responses is a difficult task, so other practical solutions have to be applied. In particular, Robinson and Hawksford (1999) implement a solution based on calculating the half wave response of the cells, and then using a series of feedback loops to simulate the increased sensitivity of the inner hair cells to the onset of sounds. Other schemes might just convolve the signal with a spreading function, to simulate the dispersion of energy along the basilar membrane, and then convert the signal back to decibels (ITU, 2001). Independently of the method used, the basilar membrane representation is obtained at this point.

After a basilar membrane representation has been obtained for both the original audio signal *A*, and the watermarked audio signal *A'*, the perceived difference between the two has to be calculated. The difference between the signals at each frequency band has to be calculated, and then it must be determined at what level these differences will become audible for a human listener (Robinson & Hawksford, 1999). In the case of the ITU Recommendation ITU-R BS.1387, this task is done by calculating a series of model variables, such as excitation, modulation and loudness patterns, and using them as an input to an artificial neural network with one hidden layer (ITU, 2001). In the model proposed in Robinson and Hawksford (1999), this is done as a summation over time (over an interval of 20 ms) along with weighting of the signal and peak suppression.

The result of this process is an objective difference between the two signals. In the case of the ITU model, the result is given in a negative five-level impairment scale, just like the BAQ, and is known as the *objective difference grade* (ODG). For other models, the difference is given in implementation-dependant units. In both cases, a mapping or scaling function, from the model units to the ITU-R. 500 scale, must be used.

For the ITU model, this mapping could be trivial, as all that is needed is to add a value of five to the value of the ODG. However, a more precise mapping function could be developed. The ODG has a resolution of one decimal, and the model was not specifically designed for the evaluation

watermarking schemes. Given this, a nonlinear mapping (for example using a logarithmic function) could be more appropriate.

For other systems, determining such a function will depend on the particular implementation of the auditory model; nonetheless such a function should exist, as a correlation between objective and subjective measures was stated as an initial requirement. For example, in the case of Thiede and Kabot (1996), a sigmoidal mapping function is used. Furthermore, the parameters for the mapping function can be calculated using a control group consisting of widely available listening test data.

The resulting grade, in the five-level scale, is defined as the perceptibility of the audio watermark. This means that in order to estimate the perceptibility of the watermarking scheme, several test runs must be performed. Again, these test runs should embed a random mark on a cover signal, and a large and representative set of audio cover signals must be used. The perceptibility test score is finally calculated by averaging the different results obtained for each one of the individual tests.

FINAL BENCHMARK SCORE

In the previous sections, three different testing procedures have been proposed, in order to measure the fidelity, robustness and perceptibility of a watermarking scheme. Each one of these tests has resulted in several scores, some of which may be more useful than others. In this section, these scores are combined in order to obtain a final benchmarking score. As a result, fair comparison amongst competing technologies is possible, as the final watermarking scheme evaluation score is obtained.

In addition, another issue is addressed at this point: defining the specific parameters to be used for each attack while performing the robustness test. While the different attacks were explained in the sixth section, the strength at which they should be applied was not specified. As a general rule of thumb, it was just stated that these operations should be tested up to the point where noticeable distortion is introduced on the audio cover file.

As it has been previously discussed, addressing these two topics can prove to be a difficult task. Moreover, a single answer might not be appropriate for every possible watermarking application. Given this fact, one should develop and use a set of application-specific evaluation templates to overcome this restriction. In order to do so, an *evaluation template* is defined as a set of guidelines that specifies the specific parameters to be used for the different tests performed, and also denotes the relative importance of each one of the tests performed on the watermarking scheme. Two fundamental concepts have been incorporated into that of evaluation templates: evaluation profiles and application specific benchmarking.

Evaluation profiles have been proposed in Petitcolas (2000) as a method for testing different levels of robustness. Their sole purpose is to establish the set of tests and media to be used when evaluating a marking algorithm. For example, Table 4, which summarizes the robustness requirements imposed by various organizations, constitutes a general-purpose evaluation profile. More specific profiles have to be developed when evaluating more specific watermarking systems. For example, one should test a marking scheme intended for advertisement broadcast monitoring with a set of recordings similar to those that will be used in a real world situation. There is no point in testing such an algorithm with a set of high-fidelity musical recordings. Evaluation profiles are thus a part of the proposed evaluation templates.

Application specific benchmarking, in turn, is proposed in Pereira et al. (2001) and Voloshynovskiy, Pereira, Iquise and Pun (2001) and consists of averaging the results of the different tests performed to a marking scheme, using a set of weights that is specific to the intended application of the watermarking algorithm. In

other words, attacks are weighted as a function of applications (Pereira et al., 2001). In the specific case of the evaluation templates proposed in this document, two different sets of weights should be specified: those used when measuring one of the three fundamental characteristics of the algorithm (i.e., fidelity, robustness and perceptibility); and those used when combining these measures into a single benchmarking score.

After the different weights have been established, the *overall watermarking scheme score* is calculated as a simple weighted average, with the formula:

$$Score = w_f * s_f + w_r * s_r + w_p * s_p$$

where *w* represents the assigned weight for a test, *s* to the score received on a test, and the subscripts *f, r, p* denote the fidelity, robustness and perceptibility tests respectively. In turn, the values of s_f, s_r, and s_p are also determined using a weighted average for the different measures obtained on the specific subtests.

The use of an evaluation template is a simple, yet powerful idea. It allows for a fair comparison of watermarking schemes, and for ease of automated testing. After these templates have been defined, one needs only to select the intended application of the watermarking scheme that is to be evaluated, and the rest of the operations can be performed automatically. Nonetheless, time has to be devoted to the task of carefully defining the set of evaluation templates for the different applications sought to be tested. A very simple, general-purpose evaluation template is shown in Box 1, as an example.

Presenting the Results

The main result of the benchmark presented here is the overall watermarking scheme score that has just been explained. It corresponds to a single, numerical result. As a consequence, comparison between similar schemes is both quick and easy. Having such a comprehensive quality measure is sufficient in most cases.

Under some circumstances the intermediate scores might also be important, as one might want to know more about the particular characteristics of a watermarking algorithm, rather than compare it against others in a general way. For example, one might just be interested in the perceptibility score of the echo watermarking algorithm, or in the robustness against uniform noise for two different schemes. For these cases, the use of graphs, as proposed in Kutter and Hartung (2000) and Kutter and Petitcolas (1999, 2000) is recommended.

The graphs should plot the variance in two different parameters, with the remaining parameters fixed. That is, the test setup conditions should remain constant along different test runs. Finally, several test runs should be performed, and the results averaged. As a consequence, a set of variable and fixed parameters for performing the comparisons are possible, and thus several graphs can be plotted. Some of the most useful graphs, based on the discussion presented in Kutter and Petitcolas (1999), along with their corresponding variables and constants, are summarized in Table 5.

Of special interest to some watermark developers is the use of receiver operating characteristic (ROC) graphs, as they show the relation between false positives and false negatives for a given watermarking system. "They are useful for assessing the overall behavior and reliability of the watermarking scheme being tested" (Petitcolas & Anderson, 1999).

In order to understand ROC graphs, one should remember that a watermark decoder can be viewed as a system that performs two different steps: first it decides if a watermark is present on the audio signal *A'*, and then it tries to recover the embedded watermark *W*. The first step can be viewed as a form of hypothesis testing (Kutter & Hartung, 2000), where the decoder decides between the alternative hypothesis (a watermark is present),

Box 1. Evaluation template

Application:

General Purpose Audio Watermarking

Final Score Weights:

Fidelity = 1/3, Robustness = 1/3, Perceptibility = 1/3

FIDELITY TEST

Measure	Parameters	Weight
Quality	N/A	0.75
Data Payload	Watermark length = 100 bits, score calculated as BER.	0.125
Speed	Watermark length = 50 bits, score calculated as 1 if embedding time is less than 2 minutes, 0 otherwise.	0.125

ROBUSTNESS TEST

Measure	Parameters	Weight
D/A Conversion	D/A ↔ A/D twice.	1/14
Equalization	10 band graphic equalizer with the following characteristics:	1/14
Band-pass filtering	100 Hz – 6 kHz, 12dB/oct.	1/14
Time stretch and pitch change	+/- 10% compression and decompression	1/14
Codecs	AAC, MPEG-4 AAC with perceptual noise substitution, MPEG-1 Audio Layer 3, Windows Media Audio, and Twin-VQ at 128 kbps.	1/14
Noise addition	Adding white noise with constant level of 40dB lower than total averaged music power (SNR: 40dB)	1/14
Time scale modification	Pitch invariant time scaling of +/- 4%	1/14
Wow and flutter	0.5% rms, from DC to 250Hz	1/14
Echo addition	Delay = 100 milliseconds, feedback coefficient = 0.5	1/14
Down mixing	Stereo to mono, and Dolby surround	1/14
Sample rate conversion	44.1 kHz to 16 kHz	1/14
Dynamic range reduction	Threshold of 50dB, 16dB maximum compression Rate: 10 millisecond attack, 3 second recovery	1/14
Amplitude compression	16 bits to 8 bits	1/14

Freq. (Hz)	31	62	125	250	500	1k	2k	4k	8k	16k
Gain (db)	-6	+6	-6	+3	-6	+6	-6	+6	-6	+6

PERCEPTIBILITY TEST

Measure	Parameters	Weight
Watermark perceptibility	N/A	1

Table 5. Useful graphs when evaluating a specific watermarking scheme

Graph Type	Perceptual Quality	Robustness Measure	Strength of a Specific Attack	Data Payload
Robustness to an Attack	fixed	variable	variable	fixed
Perceptual Quality vs. Payload	variable	fixed	fixed	variable
Attack Strength vs. Perceptual Quality	variable	fixed	variable	fixed
ROC	fixed	fixed	fixed/variable	fixed

and the null hypothesis (the watermark is not present). Given these two options, two different errors can occur, as was stated in the third section: a false positive, and a false negative.

ROC graphs plot the true positive fraction (TPF) on the Y-axis, and the false positive fraction (FPF) on the X-axis. The TPF is defined by the formula:

$$TPF = \frac{TP}{TP + FN}$$

where *TP* is the number of true positive test results, and *FN* is the number of false negative tests. Conversely, the FPF is defined by:

$$FPF = \frac{FP}{TN + FP}$$

where *TN* is the number of false-positive results, and *FP* the number of true negative results. An optimal detector will have a curve that goes from the bottom left corner to the top left, and then to the top right corner (Kutter & Petitcolas, 2000). Finally, it must be stated that the same number of watermarked and unwatermarked audio samples should be used for the test, although false-positive testing can be time-consuming, as was previously discussed in this document.

Automated Evaluation

The watermarking benchmark proposed here can be implemented for the automated evaluation of different watermarking schemes. In fact, this idea has been included in test design, and has motivated some key decisions, such as the use of a computational model of the ear instead of a formal listening test. Moreover, the establishment of an automated test for watermarking systems is an industry need. This assertion is derived from the following fact: to evaluate the quality of a watermarking scheme one can do one of the following three options (Petitcolas, 2000):

- Trust the watermark developer and his or her claims about watermark performance.
- Thoroughly test the scheme oneself.
- Have the watermarking scheme evaluated by a trusted third party.

Only the third option provides an objective solution to this problem, as long as the evaluation methodology and results are transparent to the public (Petitcolas et al., 2001). This means that anybody should be able to reproduce the results easily. As a conclusion, the industry needs to establish a trusted evaluation authority in order to objectively evaluate its watermarking products. The establishment of watermark certification programs has been proposed, and projects such as the Certimark and StirMark benchmarks are

under development (Certimark, 2001; Kutter & Petitcolas, 2000; Pereira et al., 2001; Petitcolas et al., 2001). However, these programs seem to be aimed mainly at testing of image watermarking systems (Meerwald & Pereira, 2002). A similar initiative for audio watermark testing has yet to be proposed.

Nonetheless, one problem remains unsolved: watermarking scheme developers may not be willing to give the source code for their embedding and recovery systems to a testing authority. If this is the situation, then both watermark embedding and recovery processes must be performed at the developer's side, while the rest of the operations can be performed by the watermark tester. The problem with this scheme is that the watermark developer could cheat and always report the watermark as being recovered by the detector. Even if a basic zero knowledge protocol is used in the testing procedure, the developer can cheat, as he or she will have access to both the original audio file A and the modified, watermarked file $\tilde{A}$ that has been previously processed by the tester. The cheat is possible because the developer can estimate the value of the watermarked file A', even if it has always been kept secured by the tester (Petitcolas, 2000), and then try to extract the mark from this estimated signal. Given this fact, one partial solution consists of giving the watermark decoder to the evaluator, while the developer maintains control over the watermark embedder, or vice versa[19].

Hopefully, as the need for thorough testing of watermarking systems increases, watermark developers will be more willing to give out access to their systems for thorough evaluation. Furthermore, if a common testing interface is agreed upon by watermark developers, then they will not need to release the source code for their products; a compiled library will be enough for practical testing of the implemented scheme if it follows a previously defined set of design guidelines. Nonetheless, it is uncertain if both the watermarking industry and community will undergo such an effort.

CONCLUSION

Digital watermarking schemes can prove to be a valuable technique for copyright control of digital material. Different applications and properties of digital watermarks have been reviewed in this chapter, specifically as they apply to digital audio. However, a problem arises as different claims are made about the quality of the watermarking schemes being developed; every developer measures the quality of their respective schemes using a different set of procedures and metrics, making it impossible to perform objective comparisons among their products.

As the problem just described can affect the credibility of watermarking system developers, as well as the acceptance of this emerging technology by content owners, this document has presented a practical test for measuring the quality of digital audio watermarking techniques. The implementation and further development of such a test can prove to be beneficial not only to the industry, but also to the growing field of researchers currently working on the subject.

Nonetheless, several problems arise while implementing a widely accepted benchmark for watermarking schemes. Most of these problems have been presented in this document, but others have not been thoroughly discussed. One of these problems consists of including the growing number of attacks against marking systems that are proposed every year. These attacks get more complex and thus their implementation becomes more difficult (Meerwald & Pereira, 2002; Voloshynovskiy, Pereira, Pun et al., 2001); nonetheless, they need to be implemented and included if real world testing is sought.

Another problem arises when other aspects of the systems are to be evaluated. For example, user interfaces can be very important in determining whether a watermarking product will be widely accepted (Craver et al., 2000). Its evaluation is not directly related to the architecture and performance of a marking system, but it certainly will have an impact on its acceptance.

Legal constraints can also affect watermark testing, as patents might protect some of the techniques used for watermark evaluation. In other situations, the use of certain watermarking schemes in court as acceptable proofs of ownership cannot be guaranteed, and a case-by-case study must be performed (Craver, Yeo et al., 1998; Lai & Buonaiuti, 2000). Such legal attacks depend on many factors, such as the economic power of the disputing parties.

While these difficulties are important, they should not be considered severe and must not undermine the importance of implementing a widely accepted benchmarking for audio watermarking systems. Instead, they show the need for further development of the current testing techniques. The industry has seen that ambiguous requirements and unmethodical testing can prove to be a disaster, as they can lead to the development of unreliable systems (Craver et al., 2001).

Finally, the importance of a specific benchmark for audio watermarking must be stated. Most of the available literature on watermarking relates to the specific field of image watermarking. In a similar way, the development of testing techniques for watermarking has focused on the marking of digital images. Benchmarks currently being developed, such as Stirmark and Certimark, will be extended in the future to manage digital audio content (Certimark, 2001; Kutter & Petitcolas, 2000); however, this might not be an easy task, as the metrics used in these benchmarks have been optimized for the evaluation of image watermarking techniques. It is in this aspect that the test proposed in this document proves to be valuable, as it proposes the use of a psychoacoustical model in order to measure the perceptual quality of audio watermarking schemes. Other aspects, such as the use of a communications model as the base for the test design, are novel as well, and hopefully will be incorporated into the watermark benchmarking initiatives currently under development.

REFERENCES

Acken, J.M. (1998, July). How watermarking adds value to digital content. *Communications of the ACM, 41,* 75-77.

Arnold, M. (2000). Audio watermarking: Features, applications and algorithms. Paper presented at the *IEEE International Conference on Multimedia and Expo 2000.*

Arnold, M., & Schilz, K. (2002, January). Quality evaluation of watermarked audio tracks. Paper presented at the *Proceedings of the SPIE, Security and Watermarking of Multimedia Contents IV,* San Jose, CA.

Bassia, P., & Pitas, I. (1998, August). Robust audio watermarking in the time domain. Paper presented at the *9th European Signal Processing Conference (EUSIPCO'98),* Island of Rhodes, Greece.

Bender, W., Gruhl, D., Morimoto, N., & Lu, A. (1996). Techniques for data hiding. *IBM Systems Journal, 35*(5).

Boney, L., Tewfik, A.H., & Hamdy, K.N. (1996, June). Digital watermarks for audio signals. Paper presented at the *IEEE International Conference on Multimedia Computing and Systems,* Hiroshima, Japan.

Certimark. (2001). *Certimark benchmark, metrics & parameters* (D22). Geneva, Switzerland.

Chen, B. (2000). *Design and analysis of digital watermarking, information embedding, and data hiding systems.* MIT, Boston.

Chen, B., & Wornell, G.W. (1999, January). Dither modulation: A new approach to digital watermarking and information embedding. Paper presented at the *SPIE: Security and Watermarking of Multimedia Contents,* San Jose, CA.

Chen, B., & Wornell, G.W. (2000, June). Quantization index modulation: A class of provably good methods for digital watermarking and information

embedding. Paper presented at the *International Symposium on Information Theory ISIT-2000*, Sorrento, Italy.

Cox, I.J., Miller, M.L., & Bloom, J.A. (2000, March). Watermarking applications and their properties. Paper presented at the *International Conference on Information Technology: Coding and Computing, ITCC 2000*, Las Vegas, NV.

Cox, I.J., Miller, M.L., & Bloom, J.A. (2002). *Digital watermarking* (1st ed.). San Francisco: Morgan Kaufmann.

Cox, I.J., Miller, M.L., Linnartz, J.-P.M.G., & Kalker, T. (1999). A review of watermarking principles and practices. In K.K. Parhi & T. Nishitani (Eds.), *Digital signal processing in multimedia systems* (pp. 461-485). Marcell Dekker.

Craver, S., Memon, N., Yeo, B.-L., & Yeung, M.M. (1998). Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications. *IEEE Journal on Selected Areas in Communications, 16*(4), 573-586.

Craver, S., Memon, N., & Yeung, M.M. (1996). *Can invisible watermarks resolve rightful ownerships?* (RC 20509). IBM Research.

Craver, S., Perrig, A., & Petitcolas, F.A.P. (2000). Robustness of copyright marking systems. In F.A.P. Petitcolas & S. Katzenbeisser (Eds.), *Information hiding: Techniques for steganography and digital watermarking* (1st ed., pp. 149-174). Boston, MA: Artech House.

Craver, S., Wu, M., Liu, B., Stubblefield, A., Swartzlander, B., Wallach, D.S., Dean, D., & Felten, E.W. (2001, August). Reading between the lines: Lessons from the SDMI challenge. Paper presented at the *USENIX Security Symposium*, Washington, DC.

Craver, S., Yeo, B.-L., & Yeung, M.M. (1998, July). Technical trials and legal tribulations. *Communications of the ACM, 41,* 45-54.

Czerwinski, S., Fromm, R., & Hodes, T. (1999). *Digital music distribution and audio watermarking* (IS 219). University of California - Berkeley.

Dannenberg, R., & Mazzoni, D. (2002). *Audacity* (Version 0.98). Pittsburgh, PA.

Dugelay, J.-L., & Roche, S. (2000). A survey of current watermarking techniques. In F. A.P. Petitcolas & S. Katzenbeisser (Eds.), *Information hiding: Techniques for steganography and digital watermarking* (1st ed., pp. 121-148). Boston, MA: Artech House.

Gordy, J.D., & Burton, L.T. (2000, August). Performance evaluation of digital audio watermarking algorithms. Paper presented at the *43rd Midwest Symposium on Circuits and Systems*, Lansing, MI.

Initiative, S.D.M. (2000). *Call for proposals for Phase II screening technology, Version 1.0: Secure Digital Music Initiative.*

ITU. (2001). *Method for objective measurements of perceived audio quality* (ITU-R BS.1387). Geneva: International Telecommunication Union.

JASRAC. (2001). *Announcement of evaluation test results for "STEP 2001", International evaluation project for digital watermark technology for music.* Tokyo: Japan Society for the Rights of Authors, Composers and Publishers.

Johnson, N.F., Duric, Z., & Jajodia, S. (2001). *Information hiding: Steganography and watermarking - Attacks and countermeasures* (1st ed.). Boston: Kluwer Academic Publishers.

Johnson, N.F., & Katzenbeisser, S.C. (2000). A survey of steganographic techniques. In F.A.P. Petitcolas & S. Katzenbeisser (Eds.), *Information hiding: Techniques for steganography and digital watermarking* (1st ed., pp. 43-78). Boston, MA: Artech House.

Katzenbeisser, S., & Veith, H. (2002, January). Securing symmetric watermarking schemes against protocol attacks. Paper presented at the *Proceedings of the SPIE, Security and Watermarking of Multimedia Contents IV,* San Jose, CA.

Katzenbeisser, S.C. (2000). Principles of steganography. In F.A.P. Petitcolas & S. Katzenbeisser (Eds.), *Information hiding: Techniques for steganography and digital watermarking* (1st ed., pp. 17-41). Boston, MA: Artech House.

Kirovski, D., & Malvar, H. (2001, April). Robust cover communication over a public audio channel using spread spectrum. Paper presented at the *Information Hiding Workshop,* Pittsburgh, PA.

Kuczynski, A. (2000, January 6). Radio squeezes empty air space for profit. *The New York Times.*

Kutter, M., & Hartung, F. (2000). Introduction to watermarking techniques. In F.A.P. Petitcolas & S. Katzenbeisser (Eds.), *Information hiding: Techniques for steganography and digital watermarking* (1st ed., pp. 97-120). Boston, MA: Artech House.

Kutter, M., & Petitcolas, F.A.P. (1999, January). A fair benchmark for image watermarking systems. Paper presented at the *Electronic Imaging '99. Security and Watermarking of Multimedia Contents,* San Jose, CA.

Kutter, M., & Petitcolas, F.A.P. (2000). Fair evaluation methods for image watermarking systems. *Journal of Electronic Imaging, 9*(4), 445-455.

Lai, S., & Buonaiuti, F.M. (2000). Copyright on the Internet and watermarking. In F.A. P. Petitcolas & S. Katzenbeisser (Eds.), *Information hiding: Techniques for steganography and digital watermarking* (1st ed., pp. 191-213). Boston, MA: Artech House.

Meerwald, P., & Pereira, S. (2002, January). Attacks, applications, and evaluation of known watermarking algorithms with Checkmark. Paper presented at the *Proceedings of the SPIE, Security and Watermarking of Multimedia Contents IV,* San Jose, CA.

Memon, N., & Wong, P.W. (1998, July). Protecting digital media content. *Communications of the ACM, 41,* 35-43.

Mintzer, F., Braudaway, G.W., & Bell, A.E. (1998, July). Opportunities for watermarking standards. *Communications of the ACM, 41,* 57-64.

Mintzer, F., Magerlein, K.A., & Braudaway, G.W. (1996). *Color correct digital watermarking of images.*

Pereira, S., Voloshynovskiy, S., Madueño, M., Marchand-Maillet, S., & Pun, T. (2001, April). Second generation benchmarking and application oriented evaluation. Paper presented at the *Information Hiding Workshop,* Pittsburgh, PA.

Petitcolas, F.A.P. (2000). Watermarking schemes evaluation. *IEEE Signal Processing, 17*(5), 58-64.

Petitcolas, F.A.P., & Anderson, R.J. (1998, September). Weaknesses of copyright marking systems. Paper presented at the *Multimedia and Security Workshop at the 6th ACM International Multimedia Conference,* Bristol UK.

Petitcolas, F.A.P., & Anderson, R.J. (1999, June). Evaluation of copyright marking systems. Paper presented at the *IEEE Multimedia Systems,* Florence, Italy.

Petitcolas, F.A.P., Anderson, R.J., & G., K.M. (1998, April). Attacks on copyright marking systems. Paper presented at the *Second Workshop on Information Hiding,* Portland, OR.

Petitcolas, F.A.P., Anderson, R.J., & G., K. M. (1999, July). Information hiding – A survey. Paper presented at the *IEEE.*

Petitcolas, F.A.P., Steinebach, M., Raynal, F., Dittmann, J., Fontaine, C., & Fatès, N. (2001, January 22-26). A public automated Web-based evaluation service for watermarking schemes: StirMark

Benchmark. Paper presented at the *Electronic Imaging 2001, Security and Watermarking of Multimedia Contents,* San Jose, CA.

Piron, L., Arnold, M., Kutter, M., Funk, W., Boucqueau, J.M., & Craven, F. (1999, January). OCTALIS benchmarking: Comparison of four watermarking techniques. Paper presented at the *Proceedings of SPIE: Security and Watermarking of Multimedia Contents,* San Jose, CA.

RLE. (1999). *Leaving a mark without a trace* [RLE Currents 11(2)]. Available online: http://rleweb.mit.edu/Publications/currents/cur11-1/11-1watermark. htm.

Robinson, D.J.M. (2002). *Perceptual model for assessment of coded audio.* University of Essex, Essex.

Robinson, D.J.M., & Hawksford, M.J. (1999, September). Time-domain auditory model for the assessment of high-quality coded audio. Paper presented at the *107th Conference of the Audio Engineering Society,* New York.

Secure Digital Music Initiative. (2000). *Call for proposal for Phase II screening technology* (FRWG 000224-01).

Swanson, M.D., Zhu, B., Tewfik, A.H., & Boney, L. (1998). Robust audio watermarking using perceptual masking. *Signal Processing, 66*(3), 337-355.

Thiede, T., & Kabot, E. (1996). A new perceptual quality measure for bit rate reduced audio. Paper presented at the *100th AES Convention,* Copenhagen, Denmark.

Thiede, T., Treurniet, W.C., Bitto, R., Sporer, T., Brandenburg, K., Schmidmer, C., Keyhl, K., G., B. J., Colomes, C., Stoll, G., & Feiten, B. (1998). PEAQ - der künftige ITU-Standard zur objektiven messung der wahrgenommenen audioqualität. Paper presented at the *Tonmeistertagung Karlsruhe,* Munich, Germany.

Voloshynovskiy, S., Pereira, S., Iquise, V., & Pun, T. (2001, June). Attack modelling: Towards a second generation benchmark. Paper presented at the *Signal Processing.*

Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J.J., & Su, J.K. (2001, August). Attacks on digital watermarks: Classification, estimation-based attacks and benchmarks. *IEEE Communications Magazine, 39,* 118-127.

Yeung, M.M. (1998, July). Digital watermarking. *Communications of the ACM, 41,* 31-33.

Zhao, J., Koch, E., & Luo, C. (1998, July). In business today and tomorrow. *Communications of the ACM, 41,* 67-72.

ENDNOTES

1 It must be stated that when information is digital there is no difference between an original and a bit by bit copy. This constitutes the core of the threat to art works, such as music recordings, as any copy has the same quality as the original. This problem did not exist with technologies such as cassette recorders, since the fidelity of a second-generation copy was not high enough to consider the technology a threat.

2 A test subject is defined as a specific implementation of a watermarking algorithm, based on one of the general techniques presented in this document.

3 It is implied that the transmission of a watermark is considered a communication process, where the content creator embeds a watermark into a work, which acts as a channel. The watermark is meant to be recovered later by a receiver, but there is no guarantee that the recovery will be successful, as the channel is prone to some tampering. This assumption will be further explained later in the document.

[4] Or a copy of such, given the digital nature of the medium.

[5] A cover is the same thing as a work. C, the set of all possible covers (or all possible works), is known as content.

[6] This pattern is also known as a pseudo-noise (PN) sequence. Even though the watermark message and the PN-sequence are different, it is the latter one we refer to as the watermark W.

[7] The fingerprinting mechanism implemented by the DiVX, where each player had an embedder rather than a decoder, constitutes an interesting and uncommon case.

[8] This in accordance to Kerckhoff's principle.

[9] In the case of an audio recording, the symbol ℗ along with the owner name must be printed on the surface of the physical media.

[10] The registration fee at the Office of Copyrights and Patents can be found online at: http://www.loc.gov/copyright.

[11] In fact, the call for proposal for Phase II of SDMI requires this functionality (Initiative, 2000).

[12] This is very similar to the use of serial numbers in software packages.

[13] Some of the literature refers to this distortion as beating.

[14] This is known as an oracle attack.

[15] $C(x) = E\left((x-\overline{x})(x-\overline{x})^{*}\right)$

[16] THD is the amount of undesirable harmonics present in an output audio signal, expressed as a percentage. The lower the percentage the better.

[17] A description of the equipment used on a formal listening test can be found in Arnold and Schilz (2002).

[18] 1 Bark corresponds to 100 Hz, and 24 Bark correspond to 15000 Hz.

[19] This decision will be motivated by the economics of the system; that is, by what part of the systems is considered more valuable by the developer.

This work was previously published in Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property, edited by C.-S. Lu, pp. 75-125, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.12
Digital Watermarking for Multimedia Transaction Tracking

Dan Yu
Nanyang Technological University, Singapore

Farook Sattar
Nanyang Technological University, Singapore

ABSTRACT

This chapter focuses on the issue of transaction tracking in multimedia distribution applications through digital watermarking terminology. The existing watermarking schemes are summarized and their assumptions as well as the limitations for tracking are analyzed. In particular, an Independent Component Analysis (ICA)-based watermarking scheme is proposed, which can overcome the problems of the existing watermarking schemes. Multiple watermarking technique is exploited—one watermark to identify the rightful owner of the work and the other one to identify the legal user of a copy of the work. In the absence of original data, watermark, embedding locations and strengths, the ICA-based watermarking scheme is introduced for efficient watermark extraction with some side information. The robustness of the proposed scheme against some common signal-processing attacks as well as the related future work are also presented. Finally, some challenging issues in multimedia transaction tracking through digital watermarking are discussed.

INTRODUCTION

We are now in a digital information age. Digital information technology has changed our society as well as our lives. The information revolution takes place in the following two forms

- Data/information retrieval/representation
- Data/information dissemination/communication

Digital presentation of data allows information recorded in a digital format, and thus, it brings easy access to generate and replicate the information. It is such an easy access that provides the novelty

in the current phase of the information revolution. Digital technology allows primarily use with the new physical communications media, such as satellite and fiber-optic transmission. Therefore, PCs, e-mail, MPCs, LANs, WANs, MANs, intranets, and the Internet have been evolving rapidly since the 1980s. The Internet has a worldwide broadcasting capability, a mechanism for information distribution, and a medium for collaboration and interaction between individuals and their computers regardless of geographic location. This allows researchers and professionals to share relevant data and information with each other.

As image, audio, video, and other works become available in digital form, perfect copies can be easily made. The widespread use of computer networks and the global reach of the World Wide Web have added substantially an astonishing abundance of information in digital form, as well as offering unprecedented ease of access to it. Creating, publishing, distributing, using, and reusing information have become many times easier and faster in the past decade. The good news is the enrichment that this explosive growth in information brings to society as a whole. The bad news is that it can also bring to those who take advantage of the properties of digital information and the Web to copy, distribute, and use information illegally. The Web is an information resource of extraordinary size and depth, yet it is also an information reproduction and dissemination facility of great demand and capability. Therefore, there is currently a significant amount of research in intellectual property protection issues involving multimedia content distribution via the Internet.

Thus the objective of this chapter is to present multimedia transaction tracking through digital watermarking terminology. The Independent Component Analysis (ICA) technique is employed efficiently for watermark extraction in order to verify the recipient of the distributed content, and hence, to trace illegal transaction of the work to be protected.

MULTIMEDIA DISTRIBUTION FRAMEWORK THROUGH DIGITAL WATERMARKING

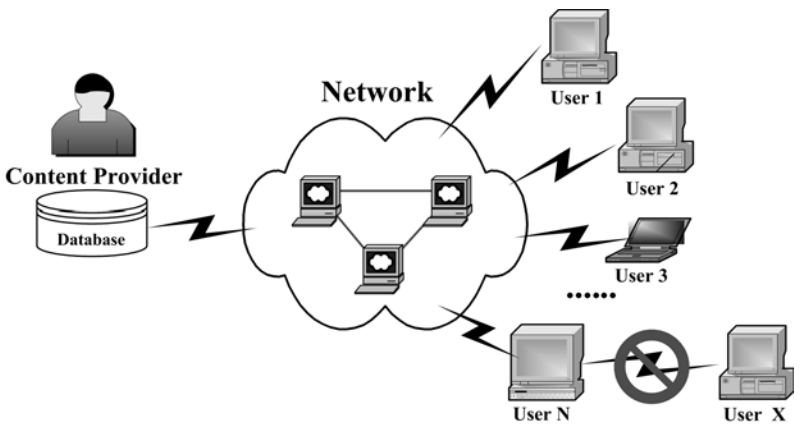
The rapid growth of networked multimedia systems has increased the need for the protection and enforcement of intellectual property (IP) rights of digital media. IP protection is becoming increasingly important nowadays. The tasks to achieve IP protection for multimedia distribution on the Internet can be classified as follows:

- **Ownership identification**: The owner of the original work must be able to provide the trustful proof that he/she is the rightful owner of the content.
- **Transaction tracking**: The owner must be able to track the distributions of the work, so that he/she is able to find the person who is responsible for the illegal replication and redistribution.
- **Content authentication**: The owner should be able to detect any illegal attempts to alter the work.

This chapter concentrates on the task of transaction tracking for multimedia distribution applications. Let us consider the scenario when an owner wants to sell or distribute the work to registered users only. To enforce IP rights, two primary problems have to be solved. First of all, the owner must be able to prove that he/she is the legal owner of the distributed content. Second, if the data have been subsequently copied and redistributed illegally, the owner must be able to find the person who is responsible for the illegal copying and redistribution of the data (see Figure 1).

The first technology adopted to enforce protection of IP rights is cryptography. Cryptographic technology (Schneier, 1995) provides an effective tool to secure the distribution process and control the legal uses of the contents that have been received by a user. The contents to be delivered

Figure 1. A multimedia distribution system where the digital content could be illegally redistributed to an illegal user



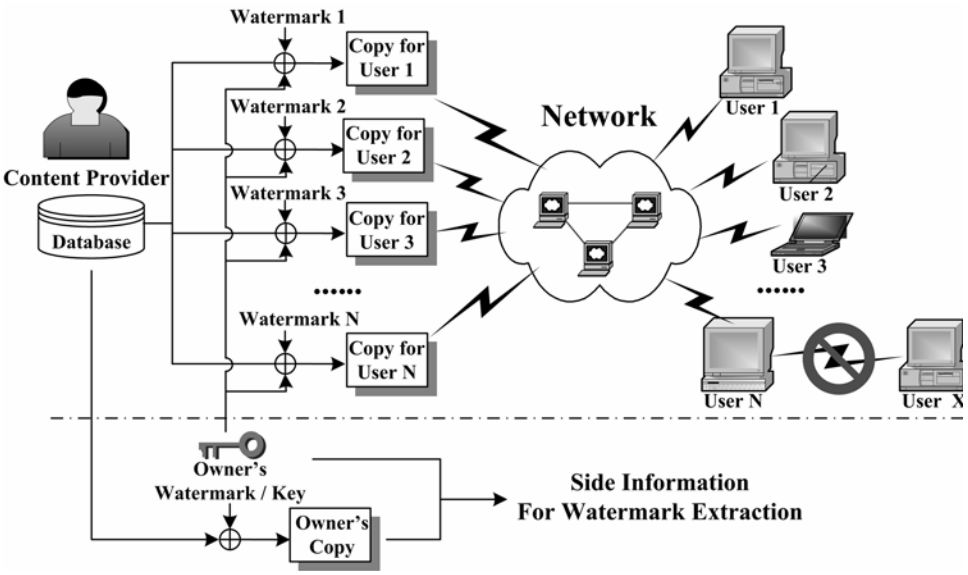
over the Internet are encrypted, and only legal users who hold the decryption key are able to use the encrypted data, whereas the data stream would be useless to a pirate without the appropriate decryption key. However, for an error-free transmission through a network, the contents after the decryption in the cryptography will be exactly the same as the original data. The data contents can be replicated perfectly many times and the user can also manipulate the contents.

Researchers and scientists are then turned to search for other technologies to counter copyright piracy on global networks that are not solvable by cryptography. In this context, recently digital watermarking technology (Cox, Miller, & Bloom, 2002) has drawn much attention. In digital watermarking, the information is transparently embedded into the work, rather than a specific media format, such as the header of a file that could be lost during transmission or file format transformation. Digital watermarking technique thus provides an efficient means for transaction tracking of illegal copying as well as redistribution of multimedia information. For a typical transaction-tracking application, the watermark identifies the first legal recipient of the work. If it is subsequently found that the work has been illegally redistributed, the watermark can then help to identify the person who is responsible for it.

Figure 2 presents a multimedia distribution framework to enforce IP rights through a technique of multiple watermarking. Multiple watermarking (Lu & Mark, 2001; Mintzer & Braudaway, 1999; Sheppard, Safavi-Naini, & Ogunbona, 2001), as suggested by the name, refers to embedding different types of watermarks into single multimedia content to accomplish different goals. For example, one of the watermarks could be used to verify ownership, the second one is to identify the recipient, and the third one is to authenticate content integrity.

For the Internet distribution application, the users first send a request to the content provider whenever they are interested for the multimedia contents. The owner can then distribute the work by signing a watermark to a registered user to uniquely identify the recipient of the work, as shown in Figure 2. All data sent to a registered user are embedded with an assigned watermark as well as the owner's watermark, while maintaining a good perceptual visual quality of the marked content. In this presented framework, the IP rights of the distributed works are enforced from the two following aspects by employing a multiple watermarking technique:

Figure 2. The multimedia distribution framework by inserting an owner's watermark to identify the ownership of the work and a unique user's watermark to identify each unique legal user



- Every copy of the work contains the owner's watermark to identify the rightful ownership of the work.
- The owner or an authorized representative can uniquely identify the recipient or the legal user of a particular copy of the multimedia content according to the embedded user's unique watermark.

Consider the case when the owner needs to prove the rightful ownership of the content. The owner can present his/her original data (without any marks) of the work as well as his/her watermark as evidence. The two embedded watermarks, including one owner's watermark and one user's watermark, are therefore able to extract by a simple subtraction method (Cox, Miller, & Bloom, 2002; Cox, Leighton, & Shamoon, 1997). One extracted watermark, that is, the owner's watermark, is matched with the previously presented owner's watermark. The rightful ownership of the content is thus verified. It is an essential prerequisite for IP protection to embed the owner's watermark into every copy of the work to be distributed over the Internet. The more difficult and challenging task, as discussed in this chapter, is to identify the legal users efficiently in the absence of the original data, and hence to trace the transactions of a particular copy of the multimedia content. For the purpose of security in multimedia, the original data are always kept in secret and should not be known to the public during watermark extraction. In some real applications, the owner needs the authorized representatives or service providers to perform the transaction-tracking tasks. For security reasons, the owner also cannot provide the original data to those representatives. Therefore, there arises a challenging issue how to extract the user's watermark in the absence of the original data. This is the main problem in transaction tracking through digital watermarking, which has been discussed in this chapter.

THE LIMITATIONS OF THE CURRENT WATERMARKING SCHEMES AND SOLUTIONS

A wide range of watermarking algorithms has been proposed. In terms of various applications, the watermarking techniques can be classified into two categories:

1. **Robust copyright marking**: to provide evidence for proving the rightful ownership of the protected digital media
2. **Authenticate marking**: to authenticate any possible alteration in the protected digital media

Robust marks can verify the ownership of the digital data, whereas the authenticate marks are used to prove whether an object has been "touched" or manipulated. This chapter focuses on robust watermarking. Robust watermarking, as opposed to authentication marking, requires the embedded watermark to be robust against all types of attacks so that it should be able to survive against attacks before the quality of the watermarked image is drastically degraded.

The major research studies on current robust watermarking techniques include the following key technical points (Barni, Bartolini, Cappellini, & Piva, 1998; Cox et al., 1997; Delaigle, Vleeschouwer, & Macq, 1998; Hartung & Kutter, 1999; Katzenbeisser & Petitcolas, 2000; Nikolaidis, & Pitas, 1998; Parhi & Nishitani, 1999):

- The choice of a work space to perform the hiding operation, mainly a spatial domain (Nikolaidis & Pitas, 1998), or a transform domain such as full-frame Discrete Cosine Transform (full DCT) (Barni et al., 1998; Cox et al., 1997; Piva, Barni, Bartoloni, & Cappellini, 1997), block DCT (Benham, Memon, Yeo, & Yeung, 1997; Hartung & Girod, 1996; Koch & Zhao, 1995; Langelaar, Lubbe, & Lagendijk, 1997; Podilchuk & Zeng, 1997; Xia, Boncelet, & Arce, 1997), Fourier Transform (FT) (Ruanaidh, Dowling, & Boland, 1996), Mellin-Fourier (Ruanaidh & Pun, 1997, 1998), or wavelet (Dugad, Ratakonda, & Ahuja, 1998; Inoue, Miyazaki, Yamamoto, & Katsura, 1998; Kundur & Hatzinakos, 1998; Wang, Su, & Kuo, 1998; Xie & Arce, 1998)
- The study of optimal watermark embedding locations based on the human visual system (Delaigle et al., 1998; Kankanhalli, Rajmohan, & Ramakrishnan, 1998; Liu, Kong, Kong, & Liu, 2001; Voloshynovskiy, Herrigel, Baumgaertner, & Pun, 1999)
- The signal embedding techniques by addition, signal-adaptive addition, or modulation methods (Cox et al., 1997; Piva et al., 1997)
- The watermark detection and extraction algorithms either in blind (Piva et al., 1997) or nonblind manner (Cox et al., 1997)

Watermark recovery is usually more robust if its original, unwatermarked data are available. For example, a simple subtraction method (Cox et al., 1997) is used for watermark extraction at the locations where watermark is embedded. The presence of the watermark is determined by cross-correlation of the original and the recovered watermark. In Piva, Barni, Bartolini, and Cappellini's method (1997), the selected DCT coefficients, where the watermark is embedded, are directly matched with all the possible watermarks stored in the database. As a result, the original data are not required for watermark detection. However, a predefined threshold is still needed to determine the presence of the watermark.

From the viewpoint of the presence of a given watermark at the extraction or verification stage, there are two different types of watermarking systems found in the literature (Katzenbeisser & Petitcolas, 2000). The first type is to embed a specific watermark or information. Most watermarking techniques for copyright protection

belong to this watermarking category, where it is assumed that the embedded watermark is previously known. The objective of this type of watermarking scheme is to verify the existence of the previously known watermark with or without the help of this watermark. The second type refers to embedding arbitrary information, which is, for example, useful for tracking the unique receipt of a copy of the distributed data. In such scenario, the watermark embedded in an image copy is previously unknown, therefore, no prior information regarding embedded watermark is available for watermark extraction. It makes the transaction tracking more difficult.

Assumptions as well as limitations for most of the existing watermarking schemes that can cause difficulties and ineffectiveness to apply in multimedia transaction tracking are summarized in the following:

(a) In some watermarking algorithms, watermark detection and extraction requires the presence of the original content. This is not desirable since the original data should always be kept secret and should not be shown to the public, or sometimes the original data are even not available immediately. Therefore, blind watermarking techniques are of great interest and concern nowadays.
(b) Most of the existing watermarking schemes (Cox et al., 1997; Cox et al., 2002; Katzenbeisser & Petitcolas, 2000) are based on some assumptions about watermark detection and extraction, such as the previous knowledge of watermark locations, strengths, or some threshold. However, in order to ensure the robustness and invisibility of the watermark, the optimum embedding locations as well as the embedding strengths are generally different for different images. For a large image database, it could be a disadvantage if it requires watermark locations and strengths information for detection and extraction of the watermark. As a result, a large amount of side information needs to be stored.
(c) As explained previously, Figure 2 shows a framework to prevent illegal redistribution of the data by a legal user. In such scenario, the current watermark detection and extraction algorithms requiring information of the watermark locations and strengths, or even the original watermark, could fail because no one knows which watermark exists in the received copy of the work.
(d) Moreover, the general watermark detection algorithm is based on a match filter finding the maximum correlation of the recovered watermark with the stored watermarks in the database containing the watermarks used to identify all possible users. It is a rather time-consuming and inefficient process, especially when a large database is needed for distribution among a large number of users.

In this chapter, an Independent Component Analysis (ICA)-based technique is proposed for watermark extraction (Yu, Sattar, & Ma, 2002). The proposed ICA-based blind watermarking scheme (Yu & Sattar, 2003) can overcome the problems of the current watermarking scheme for multimedia tracking as mentioned above. No a priori knowledge of watermark locations, strengths, threshold setting, or the original watermark is required for watermark extraction. This watermarking algorithm is found to be very effective in the application of legal data tracking compared to other watermarking algorithms. Therefore, by adopting this ICA-based watermarking approach, an efficient multimedia distribution framework for copyright protection can be accomplished.

A NEW ICA-BASED WATERMARKING SCHEME FOR MULTIMEDIA TRANSACTION TRACKING

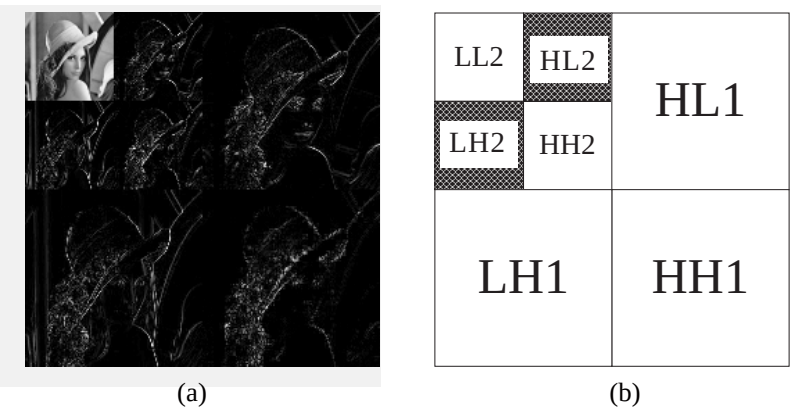
This section presents an ICA-based wavelet-domain watermarking scheme. Two watermarks are to be embedded into two selected wavelet subbands of the original image. One is the owner's watermark (or the key of the watermarking system), and the other is a unique watermark assigned to a unique legal user. The ICA technique is applied for extraction of the user's watermark with the help of side information. The proposed scheme is described in the context of watermarking in grayscale images, but this technique can be extended to color images and other digital media such as audio and video.

Proposed Watermark Embedding Scheme

Figure 3 shows a second-level wavelet decomposition of the Lena image into four bands—low-frequency band (LL), high-frequency band (HH), low-high frequency band (LH), and high-low frequency band (HL). Subbands LL and HH are not suitable for watermark embedding among these four subbands. The image quality can be degraded if the watermark is embedding in LL subband since it contains the most important information of an image. Subband HH is insignificant compared to LH and HL subbands, and watermark embedding in such subband find it difficult to survive attacks, such as lossy JPEG compression. Watermark embedding in the two subbands (e.g., LH2 and HL2 of the second-level wavelet decomposition) consisting the middle-frequency pair is to be demonstrated.

Some digital signature/pattern or company logo (**S**), for example, a text image in Figure 4(b), can be used to generate the watermark (**W**) to be embedded. This type of recognizable image pattern is more intuitive and unique than the random sequence to identify the ownership of the work. By using grayscale watermark, our method is found to be more robust against various attacks because the grayscale images can always preserve a certain level of structural information, which are meaningful and recognizable and also can be much more easily verified by human eyes rather than some objective similarity measurements. A masking function—Noise Visibility Function (NVF) (Voloshynovskiy et al., 1999)—is applied to characterize the local image properties, identifying the textured and edge regions where

Figure 3. Second-level wavelet decomposition of the Lena image



(a) (b)

Figure 4. (a) An NVF masking function, (b) a text signature (64×64 pixels), (c) the modified text watermark based on the visual mask shown in (a), (d) an owner's watermark or key of watermarking system, (e) the modified key based on (a), (f) original Lena image (256×256 pixels), and (g) a watermarked Lena image (PSNR = 45.50dB)

(a) (b) (c) (d) (e)

(f)

(g)

the information can be more strongly embedded. Such high-activity regions are generally highly insensitive to distortion. With the visual mask, the watermark strength can be reasonably increased without visually degrading the image quality.

In the next section, the watermark generation and the detailed embedding algorithm are demonstrated, followed by the generation of side information for watermark extraction.

Watermark Embedding Algorithm

Figure 5 illustrates the watermark embedding procedure using second-level decomposed middle-frequency pair (LH2 and HL2):

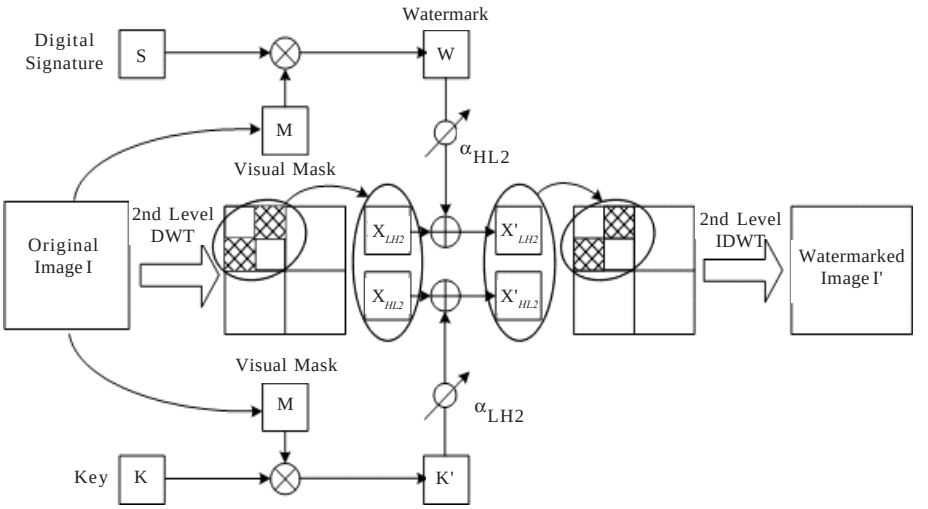
Step 1: Perform the second-level discrete wavelet decomposition of the original image **I**. Subbands LH2 and HL2 are selected for the watermark insertion.

Step 2: The NVF masking function (Voloshynovskiy et al., 1999), **M**, of the original image is generated. Figure 4(a) shows an NVF mask for the Lena image. For the homogeneous region, NVF approaches 1 (white color), and the strength of the embedded watermark approaches 0. The watermark should be embedded in highly textured regions containing edges instead of homogeneous regions. The original signature image, **S**, is modified according to the generated NVF masking function to ensure the imperceptibility of the watermark embedded. The final watermark is quantized into 0–7 gray levels. The expression for watermark generation is given as

$$\boldsymbol{W} = \boldsymbol{Q}_8 [(1 - \boldsymbol{M}) . \boldsymbol{S}], \qquad (1)$$

where $\mathbf{Q}_8$ denotes the quantization operator with 8 gray levels. Figure 4(c) shows a text watermark generated using the NVF masking function shown in Figure 4(a).

Figure 5. The proposed watermark embedding algorithm (for second-level wavelet decomposition)



Step 3: The key **K,** which is also the owner's watermark, is preprocessed by multiplying the same visual mask **M** as

$$\boldsymbol{K'} = \boldsymbol{Q}_8 [(1 - \boldsymbol{M}) . \boldsymbol{K}], \qquad (2)$$

where $\mathbf{Q}_8$ denotes the quantization operator with 8 gray levels. Figure 4(d) gives a key image for ownership authentication. Figure 4(e) shows the modified key after preprocessing by using the NVF masking function in Figure 4(a).

Step 4: The watermark **W** and the modified key **K′**are inserted into the LH2 and HL2 subband, respectively, in the following way:

$$\begin{aligned} \mathbf{X'}_{LH2} &= \mathbf{X}_{LH2} + \alpha_{LH2} \cdot \mathbf{W} = \mathbf{X}_{LH2} + \alpha_{\mathbf{x}} \cdot \mu(|\mathbf{X}_{LH2}|) \cdot \mathbf{W}; \\ \mathbf{X'}_{HL2} &= \mathbf{X}_{HL2} + \alpha_{HL2} \cdot \mathbf{K}2 = \mathbf{X}_{HL2} + \alpha_{\mathbf{x}} \cdot \mu(|\mathbf{X}_{HL2}|) \cdot \mathbf{K'}, \end{aligned} \qquad (3)$$

where **X** and **X′**are the wavelet transform coefficients of the original and the watermarked image, respectively. In Equation 3, α_{LH2} and α_{HL2} denote the weighting coefficients of the watermark embedding in subbands LH2 and HL2, respectively, while μ (| |) denotes the mean of the absolute value. A common control parameter $\alpha_{\mathbf{x}}$ in Equation 3 is used to adjust the watermark embedding strength to preserve a satisfactory quality of the final watermarked image (Peter, 2001).

Step 5: The watermarked image **I′** is obtained by the inverse discrete wavelet transform.

Step 6: Steps 4 and 5 are repeated until the quality of the final watermarked image is satisfactory, for instance, the PSNR (peak signal-to-noise ratio) measure is within the range of 40–50dB. Particularly the parameter $\alpha_{\mathbf{x}}$ is tuned to adjust the watermark strength to obtain the desired embedding result. Decreasing the magnitude of $\alpha_{\mathbf{x}}$ results in a better quality of the final marked image and vice versa. Figure 4(e) shows a watermarked Lena image with PSNR 45.50dB. Table 1 shows the quality of watermarked image (in dB) with respect to the control parameter $\alpha_{\mathbf{x}}$.

Side Information for Watermark Extraction

As discussed earlier, the original data may be unavailable in many real applications for security

Table 1. PSNR (in dB) of the watermarked image with respect to α_x

α_x	0.01	0.05	0.10	0.15	0.20	0.25	0.30
PSNR (dB)	67.50	53.52	47.50	43.98	41.48	39.54	37.96

purposes. In order to identify legal users, some side information is necessary to extract the users' watermarks in the absence of the original data. The proposed method allows the owner to create a copy of the data set by embedding only the owner's watermark following the same procedure shown in Figure 5. The owner's watermark is, in fact, the key of the watermarking system that is used for watermark extraction. Using only the owner's copy $\mathbf{I}'_0$ and the key $\mathbf{K}$, the owner is able to identify the recipient of any distributed image by ICA methods. This will be elaborated in the next subsection.

Figure 6 illustrates an owner's copy of the Lena image, embedded with the owner's watermark shown in Figure 4(d). The owner's copy is then obtained by embedding the modified key $\mathbf{K}'$ in the wavelet domain as follows:

$$\mathbf{X}'_{0LH2} = \mathbf{X}_{LH2} + \alpha_{LH2} \cdot \mathbf{K}' = \mathbf{X}_{LH2} + \alpha_{X0} \cdot \mu(|\mathbf{X}_{0LH2}|) \cdot \mathbf{K}';$$
$$\mathbf{X}'_{0HL2} = \mathbf{X}_{HL2} + \alpha_{HL2} \cdot \mathbf{K}' = \mathbf{X}_{HL2} + \alpha_{X0} \cdot \mu(|\mathbf{X}_{0HL2}|) \cdot \mathbf{K}', \quad (4)$$

Figure 6. The owner's copy of the Lena image (256×256 pixels and PSNR = 46.72dB)

where $\mathbf{X}_0$ and $\mathbf{X}'_0$ are respectively the wavelet transform coefficients of the original image and the watermarked channel, and α_{X0} is a control parameter for the visual quality of the watermarked image $\mathbf{I}'_0$.

Suppose an owner wants to authorize a third party, called appointed representative, to do the tracing task. In such case, the owner should also assign a unique watermark to the appointed representative. This representative's watermark would then replace the owner's watermark embedded in the HL2 wavelet subband. It would also be used as the key during watermark extraction. However, at the same time, for ownership verification, the owner's watermark still needs to be embedded in the wavelet subband selected other than the LH2 and HL2 subbands.

Proposed Watermark Extraction Scheme Using the ICA Method

In this section, the concept of ICA is briefly introduced. Then a blind watermark extraction scheme is proposed. The ICA technique is employed for watermark extraction successfully, without knowing the original image and any prior information on the embedded watermark, embedding locations, and strengths.

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is one of the most widely used methods for performing blind source separation (BSS). It is a very general-purpose statistical technique to recover the independent sources given only sensor observations that are linear mixtures of independent source signals (Hyvärinen, 1999b; Hyvärinen & Oja, 1999; Lee, 1998). ICA has been widely applied in

many areas such as audio processing, biomedical signal processing, and telecommunications. In this paper, ICA is further applied in watermarking for blind watermark extraction.

The ICA model consists of two parts: the mixing process and unmixing process. In the mixing process (Hyvärinen, 1999b; Hyvärinen & Oja, 1999; Lee, 1998), the observed linear mixtures $x_1, \ldots, x_m$ of n number of independent components are defined as

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \ldots + a_{jn}s_n;\ 1 \le j \le m, \quad (5)$$

where $\{s_k, k = 1, \ldots, n\}$ denote the source variables, that is, the independent components, and $\{a_{jk}, j = 1, \ldots, m; k = 1, \ldots, n\}$ are the mixing coefficients. In vector-matrix form, the above mixing model can be expressed as

$$\mathbf{x} = \mathbf{As}, \quad (6)$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

is the mixing matrix (Hyvärinen, 1999b; Hyvärinen & Oja, 1999; Lee, 1998), $\mathbf{x} = [x_1\ x_2\ \ldots\ x_m]^T$, $\mathbf{s} = [s_1\ s_2\ \ldots\ s_n]^T$, and T is the transpose operator. For the unmixing process (Hyvärinen, 1999b; Hyvärinen & Oja, 1999; Lee, 1998), after estimating the matrix **A**, one can compute its inverse—the unmixing matrix **B** and the independent components are obtained as

$$\mathbf{s} = \mathbf{Bx}. \quad (7)$$

To ensure the identifiability of the ICA model, the following fundamental restrictions are imposed (Hyvärinen, 1999b; Hyvärinen & Oja, 1999):

- The source signals in the mixing process should be principally statistically independent.
- All independent components s_k, with the possible exception of one component, must be non-Gaussian.
- The number of observed linear mixtures m must be at least as large as the number of independent components n, that is, $m \ge n$.
- The matrix **A** must be of full column rank.

There are many ICA algorithms that have been proposed recently. Some popular ICA methods include Bell and Sejnowski's Infomax (1995), Hyvärinen and Oja's FastICA (1999), Cichocki and Barros' RICA (Robust batch ICA) (1999), Cardoso's JADE (Joint Approximate Diagonalization of Eigen-matrices) (1999), and so on. From the stability standpoint, it is more appropriate to choose RICA or JADE algorithms than Infomax and FastICA algorithms for our watermark extraction process. Both Infomax algorithm and FastICA algorithm require that the values of the mixing coefficients for the sources not be very close (Bell & Sejnowski, 1995; Hyvärinen, 1999a). However, both the watermark and the key are embedded by multiplication with small weighting coefficients to make them invisible. Therefore, the extraction of such weak watermark signals could fail by using Infomax or FastICA algorithm. The extraction results using FastICA algorithm also very much depend on the initial guess of the weighting coefficients (Hyvärinen, 1999a).

Cichocki and Barro's RICA algorithm is an effective blind source separation approach particularly for the temporally correlated sources, since it models the signal as an autoregressive (AR) process (Cichocki & Barros, 1999). The RICA algorithm thus can achieve the best extraction results when both the embedding and extraction are performed in the spatial domain. This is because, generally speaking, the natural images are spatially correlated and can be effectively modeled

as temporally correlated sequences. However, for the proposed watermarking scheme described in this chapter, the watermark is embedded in the wavelet domain instead of the spatial domain. The experimental results show that the JADE algorithm (Cardoso, 1999) outperforms the other ICA algorithms for watermark extraction in our proposed watermarking scheme. This could be due to the use of higher-order statistical parameters in the JADE algorithm, such as fourth-order cumulant, which can model the statistical behavior of the wavelet coefficients more effectively. Therefore, the JADE algorithm is employed to elaborate the watermark extraction process in our proposed watermarking scheme, which will be described next.

Proposed Blind Watermark Extraction Scheme

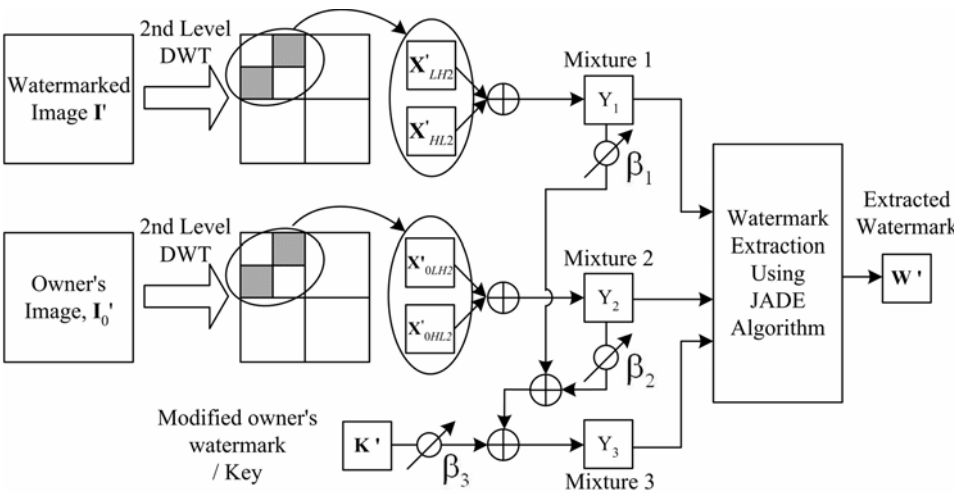
This section proposes the ICA-based blind watermark extraction scheme. Instead of using the original image, only an owner's copy of the original image is required for watermark extraction. The new useful feature of the proposed scheme is that the proposed method does not require previous knowledge of the original watermark, embedding locations, and watermark strengths for extraction. The main idea is to consider two subbands ($\mathbf{X}'_{\mathbf{R}}$) of the watermarked image to have a mixture image of the wavelet transformed image ($\mathbf{X}_{\mathbf{R}}$) of the original image (**I**), the watermark image (**W**), and the modified key (**K**′). Figure 7 shows the proposed blind watermark extraction scheme. Let us denote the received watermarked image as $\tilde{\mathbf{I}}'$. The symbol (~) is to indicate that the received data may or may not be the same as its original watermarked data due to transmission errors or possibly pirate attacks. This symbol (~) is removed in the following for simplicity.

Step 1: Perform the second-level discrete wavelet decomposition of the watermarked image **I**′ in order to obtain the wavelet coefficients $\mathbf{X}'_{LH2}$ and $\mathbf{X}'_{HL2}$ for the two selected subbands of LH2 and HL2.

Step 2: The first mixture signal $\mathbf{Y}_1$ is obtained by

$$\mathbf{Y}_1 = \mathbf{X}'_{\mathrm{LH2}} + \mathbf{X}'_{\mathrm{HL2}}. \tag{8}$$

Figure 7. Proposed blind watermark extraction scheme (using the second-level decomposition)



From Equation 3, $\mathbf{X}'_{\mathbf{R}}$ ($\mathbf{R} \in [LH2, HL2]$) are the mixture observations of the wavelet transform of the original image ($\mathbf{X}_{\mathbf{R}}$), the watermark (**W**), and the modified key (**K**′), therefore, Equation 8 can be rewritten as

$$\mathbf{Y}_1 = \mathbf{X} + \alpha_1 \mathbf{W} + \alpha_2 \mathbf{K}', \quad (9)$$

where $\mathbf{X} = \mathbf{X}_{LH2} + \mathbf{X}_{HL2}$, $\alpha_1 = \alpha_{\mathbf{X}} \cdot \mu(|\mathbf{X}_{LH2}|)$ and $\alpha_2 = \alpha_{\mathbf{X}} \cdot \mu(|\mathbf{X}_{HL2}|)$. It is found that the first mixture signal is a linear mixture of the three independent sources, that is, **X**, **W** and **K**′.

Step 3: Repeat the procedure in Steps 1 and 2 for the owner's image $\mathbf{I}'_0$. The second mixture $\mathbf{Y}_2$ is obtained by

$$\mathbf{Y}_2 = \mathbf{X}'_{0LH2} + \mathbf{X}'_{0HL2}. \quad (10)$$

Similarly $\mathbf{Y}_2$ is also a linear mixture of the wavelet transform of the original image ($\mathbf{X}_{\mathbf{R}}$, $\mathbf{R} \in [LH2, HL2]$) and the key/owner's watermark (**K**). It can be written as

$$\mathbf{Y}_2 = \mathbf{X} + \alpha_3 \mathbf{K}', \quad (11)$$

where $\alpha_3 = \alpha_{\mathbf{X}0} \cdot [\mu(|\mathbf{X}_{0LH2}|) + \mu(|\mathbf{X}_{0HL2}|)]$.

Step 4: From Equations 8 and 10, two mixture images can be obtained containing three sources or independent components in the observations—**X**, the modified key **K**′, and the watermark **W**. As was pointed out earlier, to exploit ICA methods for watermark extraction, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources in order to ensure the identifiability of the ICA model (Hyvärinen, 1999b; Hyvärinen & Oja, 1999; Lee, 1998). Therefore, another linear mixture of the three independent sources is needed. The third mixture $\mathbf{Y}_3$ can then be generated by linear superposition of $\mathbf{Y}_1$, $\mathbf{Y}_2$ and **K**′:

$$\mathbf{Y}_3 = \beta_1 \mathbf{Y}_1 + \beta_2 \mathbf{Y}_2 + \beta_3 \mathbf{K}', \quad (12)$$

Figure 8. The extraction result for the user's watermark image (normalized correlation coefficient, r = 0.9790), using JADE ICA method

where β_1 and β_2 are arbitrary real numbers, and b_3 is a nonzero arbitrary real number. Either β_1 or β_2 can be set to zero to efficiently reduce the computational load of ICA. Note that the modified key **K**′ can be easily obtained by regenerating the NVF visual mask and multiplying it to the original owner's watermark **K**.

Step 5: The three mixtures input into the JADE ICA algorithm (Cardoso, 1999) and the watermark image **W**′is extracted. The user of any image copy can be uniquely identified from the signature of the extracted watermark. Figure 8 shows the extracted watermark from the watermarked image shown in Figure 4(g).

PERFORMANCE EVALUATION

The robustness results of the proposed watermarking scheme are shown in this section using the Lena image of size 256×256 when the simulations are performed in the MATLAB 6.5 software environment. A watermarked image (PSNR = 45.50 dB) in Figure 4(g) is generated by setting the watermark strength control parameter $\alpha_{\mathbf{X}}$ as 0.15. In the experiments of watermark extraction, the parameters β_1, β_2, and β_3 are set as 0, 1, 1, respectively, to simplify the computational load of the ICA processing, and Daubechies-1 (Haar) orthogonal filters are employed for wavelet decomposition. In order to investigate the

robustness of the watermark, the watermarked image is manipulated by various signal processing techniques, such as JPEG compression and JPEG2000 compression, quantization, cropping, and geometric distortions. The watermark extraction is performed for the distorted watermarked image and the extracted watermark is compared to the original.

The Performance Index

The performance of the blind watermark extraction result is evaluated in terms of normalized correlation coefficient, *r*, of the extracted watermark $\mathbf{W}'$ and the original watermark $\mathbf{W}$ as

$$r = \frac{\mathbf{W} \cdot \mathbf{W}'}{\sqrt{\mathbf{W}^2 \cdot \mathbf{W}'^2}}. \quad (13)$$

The magnitude range of *r* is [-1, 1], and the unity holds if the matching between the extracted image and the original image is perfect.

Robustness Against Compression and Quantization Attacks

In the following, the robustness of the proposed watermarking scheme is compared with some other blind wavelet-domain watermarking schemes (Peter, 2001) in terms of normalized correlation coefficient *r* as shown in Equation 13. These techniques include Wang, Su, and Kuo's algorithm (1998), Inoue, Miyazaki, Yamamoto, and Katsura's blind algorithm (based on manipulating insignificant coefficients) (1998), Dugad, Ratakonda, and Ahuja's algorithm (1998), Kundur and Hatzinakos' algorithm (1998), and Xie and Arce's algorithm (1998).

Wang et al. (1998) have proposed an adaptive watermarking method to embed watermarks into selected significant subband wavelet coefficients. A blind watermark retrieval technique has been proposed by truncating selected significant coefficients to some specific value. Inoue et al. (1998) have classified insignificant and significant wavelet coefficients using the embedded zerotree wavelet (EZW) algorithm. Thereby, two types of embedding algorithms have been developed in respect to the locations of significant or insignificant coefficients. Information data are detected using the position of the zerotree's root and the threshold value after decomposition of the watermarked image. Dugad et al. (1998) have added the watermark in selected coefficients having significant energy in the transform domain to ensure nonerasability of the watermark. During watermark detection, all the high-pass coefficients above the threshold are chosen and are correlated with the original copy of the watermark. Kundur and Hatzinakos (1998) have presented a novel technique for the digital watermarking of still images based on the concept of multiresolution wavelet fusion, which is robust to a variety of signal distortions. Xie and Arce (1998) have developed a blind watermarking digital signature for the purpose of authentication. The signature algorithm is first implemented in the wavelet-transform domain and is coupled within the SPIHT (Set Partitioning in Hierarchical Trees) compression algorithm.

Figure 9 shows the comparison results in terms of performance index against the JPEG compression. For the proposed scheme, the extracted watermark's correlation decreases gradually with the compression quality factor. The image quality (in PSNR) has degraded significantly to 27 dB when the compression quality becomes quite low to 10%. In such a difficult case, the watermark can still be extracted with the value of *r* equal to 0.2553 for watermark embedding in second-level wavelet decomposition. According to Figure 9, the presented method can perform better than the Wang et al.'s and the Kundur and Hatzinakos' methods, while performing much better than the Inoue et al.'s method in terms of robustness against JPEG compression attack at a very low compression quality factor.

Figure 10 is the extraction comparison against the JPEG2000 compression attacks. The robustness of the proposed scheme is demonstrated up

Figure 9. Comparison of results against JPEG compression attacks

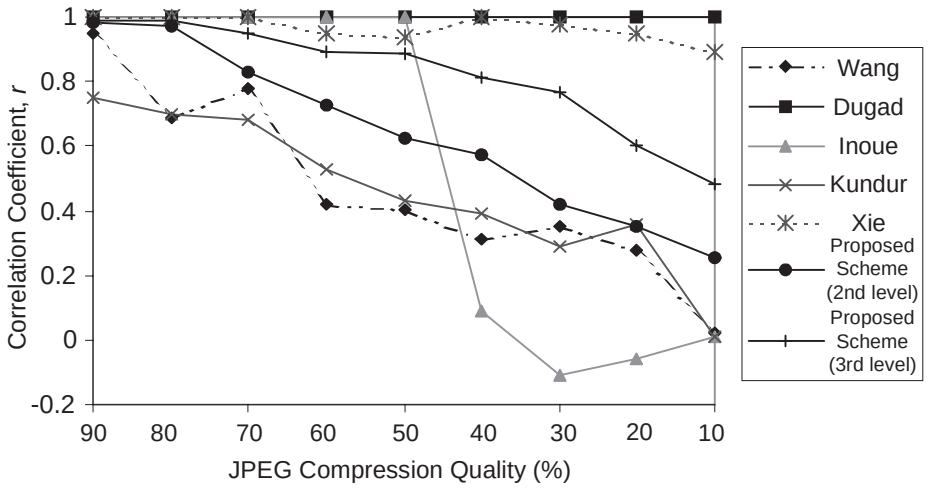
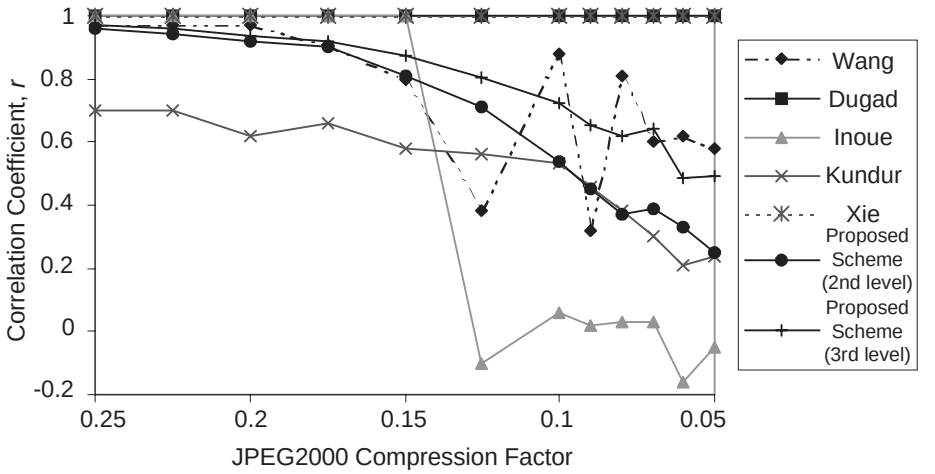


Figure 10. Comparison of results against JPEG2000 compression attacks

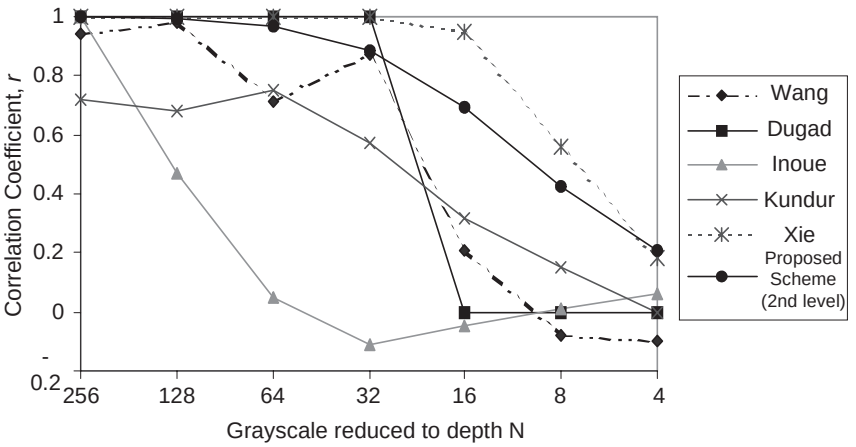


to the compression factor 0.05 or compression rate 0.4 bpp (bit per pixel). The proposed scheme gives better performance than Kundur and Hatzinakos' method, and comparable performance to the Wang et al.'s method. The extraction performance of the Inoue et al.'s method drops sharply when the JPEG2000 compression factor decreases to 0.125. Embedding in the subbands of higher wavelet decomposition level (see curves for third-level decomposition in Figures 9 and 10) can improve significantly the robustness of the proposed scheme against compression attacks.

Figure 11 shows the extraction results against quantization from gray level 256 to gray level 4 per pixel. The proposed scheme has very good robustness result against quantization. The performance of the proposed scheme is comparable to that of the Xie and Arce's method, and much better than the other methods.

From Figures 9 and 10, it is found that Xie and Arce's and Dugad et al.'s methods have excellent robustness performance against JPEG and JPEG2000 compression. In Xie and Arce's algorithm, the watermark is embedded solely in

Figure 11. Comparison of results against quantization



the approximation image (LL subband) of the host image (Xie & Arce, 1998). Although LL subband embedding is robust against compression attacks, the image quality could be degraded visually because the coefficients of this portion always contain the most important information of an image (Peter, 2001). It is claimed that the robustness of Xie and Arce's method very much depends on the number of decomposition levels. Very good robustness result can be obtained by employing a five-level wavelet decomposition using Daubechies-7/9 bi-orthogonal filters (Peter, 2001; Xie & Arce, 1998). On the other hand, in the Dugad et al.'s method, the watermark is embedded in the significant coefficients of all detail subbands (Dugad et al., 1998); therefore, it is more resistant to compression. During the watermark detection using Dugad et al.'s method, the original watermark is required to compute the correlation for the high-pass coefficients with the values above a threshold (Dugad et al., 1998). The presence of the watermark is determined by comparing this correlation with a threshold setting. It is not as general as our proposed scheme where the original watermark and the threshold are not required for watermark extraction.

The experimental results show that the proposed scheme has good robustness against the most prominent attacks such as JPEG and JPEG2000 compression, quantization, and can be comparable to existing blind wavelet-domain watermarking schemes. Experimental results also show that unlike the Xie and Arce's method (Peter, 2001; Xie & Arce, 1998), the choice of the wavelet transform is not critical concerning the robustness issue of the proposed watermarking method (the corresponding results are not included here).

Robustness Against Cropping and Geometric Distortions

Many watermarking techniques cannot survive geometric transformations such as rotation, scaling, and translation (RST) and sometimes cropping attack as well due to the loss of the synchronization of the watermark detector. A solution to such geometric attacks is to apply a resynchronization process (blind or nonblind) before performing the watermark extraction. Nonblind solution requires the presence of the original data, or at least some knowledge of the image features (Dugelay & Petitcolas, 2000). Davoine, Bas, Hébert, and Chassery (1999) have proposed a nonblind solution by splitting the original image into a set of triangular patches. This mesh serves

as a reference mesh and is kept in the memory for synchronization preprocessing. This proposed method, however, is only efficient in the case of involving minor deformations. Johnson, Duric, and Jajodia (1999) have proposed a method to invert affine transformations by estimating the difference in the least square sense between the salient image feature points in the original and transformed images. Kutter (1998) has proposed alternative methods to retrieve the watermark from geometrically distorted image without using the original data. The first method is to preset a part of the watermark to some known values and to use them for spatial resynchronization. This approach decreases the hiding capacity of the useful information, and is also computationally very expensive. The second method proposed by Kutter (1998) is to use self-reference systems that embed the watermark several times at the shifted locations.

Generally speaking, the tuning process can be easier, more accurate and requires less computational load when the original data or reference feature points are available, although it may need extra memory to store the reference data. In our proposed watermarking scheme, original data are not available during the extraction process; however, an owner's or a representative's copy of the data is available. This image copy would be very similar to the original data, thus it is convenient to use it directly as a reference for synchronization of geometric distorted or cropped data. By simple comparisons, the tampered data can be adjusted back to original size and position rapidly and accurately. In the following, the watermark extraction results against attacks of cropping and RST are shown. The effectiveness of employing synchronization preprocessing is demonstrated by showing the significant improvements of extraction results with and without the synchronization.

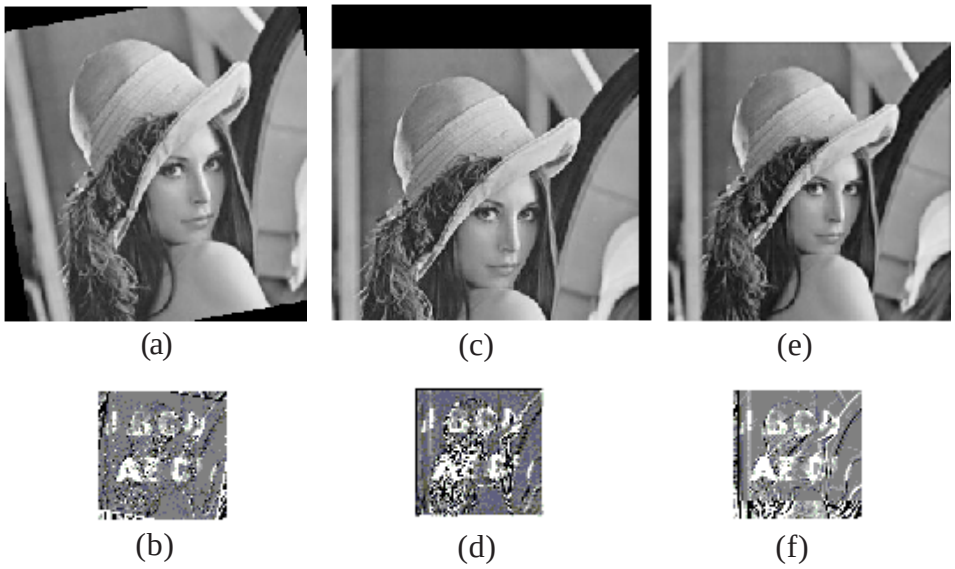
As shown in Figure 12(a), the face portion of a marked Lena image is cropped. By comparison with the owner's Lena image copy, it can be easily detected that the pixels within a square area, with row index from 121 to 180 and column index from 121 to 180, are corrupted. The absence of the watermark information in this corrupted region (by considering both rows and columns from 31 ($\lceil 121/4 \rceil$) to 45 ($\lceil 180/4 \rceil$) results in an undesired overbrightness effect for the extracted watermark due to its high values in the corrupted region. This makes both the subjective and the objective verification measurements quite poor (see Figure 12(b)). One simple solution is to discard the corresponding undesired high-valued pixels from the extracted watermark and replace them with zero-valued pixels. In this way, according to Figure 12(c), the extracted watermark can be recovered mostly with some information loss in the corrupted region. Therefore, the normalized correlation coefficient r is found to increase from 0.2706 to 0.9036, showing the recovery of the low-contrast watermark (compare Figures 12(b) and 12(c)).

Figure 12. (a) A cropped Lena image, (b) the extracted watermark (r = 0.2706), and (c) the extracted watermark after improving the contrast of (b) (r = 0.9036)

(a)

(b) (c)

Figure 13. (a) Lena image rotated by 10º, and (b) the corresponding extracted watermark (r = 0.5847); (c) Lena image downsized by reducing the number of rows and columns by 1/4, and (d) the corresponding extracted watermark (r = 0.4970); (e) Lena image translated to the left and downward by 10 and 36 lines, respectively, and (f) the corresponding extracted watermark (r = 0.5356)



The watermark extraction of the geometrically distorted image may fail due to the loss of synchronization. A resynchronization preprocessing of the received data is necessary to tune it back to the right positions or sizes before input in the watermark decoder. However, the side information in the proposed watermarking scheme—the owner's or the representative's copy of the original image—provides a good reference to assist the synchronization process, and the watermark extraction performance is consequently improved. Figures 13(b), 13(d), and 13(f) show the extraction results under attacks of rotation, scaling, and translation (RST), respectively, after the synchronization process. The watermark extraction results are satisfactory in terms of both the subjective visualization and the objective similarity measurement.

Discussions

Watermarking versus Fingerprinting

To enforce IP rights for multimedia distribution over the Internet, it requires not only verification of the legal recipient, but also proof of the rightful ownership of the original work. The term *fingerprinting* (Katzenbeisser & Petitcolas, 2000; Arnold et al., 2003; Trappe, Wu, Wang, & Liu, 2003) is closely related to watermarking in the context of traitor tracking problem. Fingerprinting technique involves the embedding of different watermarks into each distributed copy of the work. The purpose of fingerprinting is to identify the legal recipient rather than the source of digital data. Thus, using the fingerprinting technique alone is not sufficient to enforce IP rights protection in multimedia distribution systems, as the owner

cannot provide trustful evidence for proving the ownership. This chapter presents a multiple digital watermarking framework that can achieve the above two demands, that is, identifying the owner and the recipient of the distributed content.

Moreover, fingerprinting has another distinct interpretation, which does not involve the concept of digital watermarking at all. It refers to the extraction of unique features, such as semantically relevant or characteristic features from multimedia signals, in order to distinguish itself from other similar objects (Katzenbeisser & Petitcolas, 2000; Arnold et al., 2003). The extracted features are normally stored separately as signatures for authentication of the content rather than inserting them into the content as watermarks. This concept falls out of scope of this chapter.

Summary and Future Work of the Proposed Watermarking Scheme

The proposed ICA-based watermarking scheme shows its main advantage in terms of generality. Unlike other methods, no a priori information about the original watermark, embedding locations, strengths, as well as the threshold is needed for our blind watermark extraction scheme. Therefore, it is possible to extract the watermark from any copy of the watermarked image, where the embedded watermark is previously unknown. The other advantage of the proposed ICA-based method is that without using a pre-defined threshold, the extracted watermark could simply be verified from visual inspection instead of using the correlation-based matching technique with a threshold setting. This is possible because the embedded watermark used in our scheme is a readable digital signature image or a logo image. The generality of the proposed scheme implicates this method to be a quite useful tool for the transaction tracking in the application of Internet distribution. The only disadvantage to achieving the generality using ICA-based technique could be the complexity of the ICA itself. In this chapter, this has been compromised by the use of JADE algorithm, which is simple and computationally efficient. Furthermore, there are only three mixing sources (i.e., the original data, the key, and the watermark) involved in the presented watermarking scheme, which enables our ICA-based extraction processing to be fast.

In the future, more experiments need to be carried out in order to evaluate the resilience of this scheme against other types of attacks. For example, the collusion attacks and the possible counterattacks for multimedia distribution systems are to be investigated to improve the present scheme. The issue on the generation of a better perceptual mask as used to simulate the human visual system should also be studied to improve the visual quality of the watermarked data.

CONCLUDING REMARKS AND FUTURE TRENDS

This chapter investigates the multimedia transaction tracking issue by means of digital watermarking. One feasible solution of using an ICA-based watermarking technique is presented to perform ownership verification and traitor tracking for multimedia distribution through public networks. Two watermarks consisting of an owner's watermark for ownership identification and a user's watermark for unique recipient identification are embedded. Watermark is obtained by modification of the signature image with a visual mask in order to prevent the perceptual quality degradation of the watermarked image. The watermarks are inserted in the two middle frequency subband pair at the higher wavelet decomposition level (say second/third decomposition level) of the original image. Without requiring any information such as original watermark, embedding locations, and strengths, our proposed scheme can extract the user's watermark with the help of an owner's copy of the image and the owner's watermark/key. Experimental results show that the proposed wa-

termarking scheme can provide good resistance to attacks of image compression, quantization, cropping, and geometric transformations.

It has been elaborated in this chapter that the ICA-based watermarking scheme can be employed as an efficient tool to trace the recipient of the distributed content. From the general perspective, the challenging issues of digital watermarking in the applications of transaction tracking for the Internet distribution include the following criteria:

- The original data are not available during extraction of the recipient's watermark. Thus the watermarking technique should be blind.
- No prior information about the embedded watermark and the corresponding locations is available for watermark extraction.
- In order to present as trustful evidence in the court to litigate the pirate, a highly robust watermarking scheme against common signal possessing attacks as well as collusion attacks is needed.
- For some applications, for example, searching for the pirated watermarked image using Web crawlers, it is required that the watermarking scheme is able to extract the watermark easily and with low complexity.

There is no doubt that transaction tracking is a more difficult task than copyright protection by means of digital watermarking. More general watermarking techniques are desired such that no original data and prior watermark information is needed for extraction, while providing the methods to be reliable, robust, and computationally efficient.

Another requirement to enforce the IP rights of the distributed work could be such that the owner should be able to detect any illegal attempts manipulating the content. Fragile watermark should be inserted as well in order to protect the integrity of the content. The authentication watermark should be then very sensitive to various attacks and, therefore, able to locate possible modifications. In such scenario, three watermarks would be hidden in the work in order to verify the owner, to identify the user, and to authenticate the content. Since the information-hiding capacity of the cover media is limited, we have further challenges to investigate, for example, how to compromise the three demands including the information-hiding capacity and the imperceptibility and robustness of the hidden watermark.

There has been no rush yet to embrace any of the current watermarking schemes for IP rights protection in multimedia distribution application. In fact, time is still needed for thorough inspection and appraisal to find solutions for better digital watermarking schemes. Before that, scientists and researchers have to fully understand the practical requirements associated with the real problems. In the meantime, the main challenge for researchers is to develop even more transparent and decodable schemes for robust or fragile watermarking, or perhaps to meet more demands required for a practical multimedia distribution system.

ACKNOWLEDGMENT

The authors would like to thank Professor N.G. Kingsbury for his valuable suggestions and comments that helped to improve the proposed watermarking scheme for transaction tracking in multimedia distribution applications. They are also thankful to Professor K.-K. Ma for contributing useful discussion regarding the use of ICA in image watermarking.

REFERENCES

Arnold, M., Schmucker, M., & Wolthusen, S.D. (2003). *Techniques and applications of digital watermarking and content protection*. Boston: Artech House.

Barni, M., Bartolini, F., Cappellini, V., & Piva, A. (1998). A DCT-domain system for robust image watermarking. *Signal Processing, 66,* 357–372.

Barni, M., Bartolini, F., Cappellini, V., Piva, A., & Rigacci, F. (1998). A M.A.P. identification criterion for DCT-based watermarking. *Proc. Europ. Signal Processing Conf. (EUSIPCO'98),* Rhodes, Greece.

Bell, A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Compt., 7,* 1129–1159.

Benham, D., Memon, N., Yeo, B.-L., & Yeung, M. (1997). Fast watermarking of DCT-based compressed images. *Proc. Int. Conf. Image Science, System, and Technology (CISST'97),* Las Vegas, NV.

Cardoso, J.-F. (1999). High-order contrasts for independent component analysis. *Neural Computer, 11,* 157–192.

Cichocki, A., & Barros, A.K. (1999). Robust batch algorithm for sequential blind extraction of noisy biomedical signals. *Proc. ISSPA'99, 1,* 363–366.

Cox, I.J., Leighton, F.T., & Shamoon, T. (1997). Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Processing, 6,* 1673–1687.

Cox, I.J., Miller, M.L., & Bloom, J.A. (2002). *Digital watermarking.* Morgan Kaufmann.

Davoine, F., Bas, P., Hébert, P.-A., & Chassery, J.-M. (1999). Watermarking et résistance aux déformations géométriques. *Cinquièmes journées d'études et d'échanges sur la compression et la représentation des signaux audiovisuels (CORESA'99),* Sophia-Antiplis, France.

Delaigle, J.F., Vleeschouwer, C.D., & Macq, B. (1998). Watermarking algorithm based on a human visual model. *Signal Processing, 66,* 319–335.

Dugad, R., Ratakonda, K., & Ahuja, N. (1998). A new wavelet-based scheme for watermarking images. *Proc. Int. Conf. Image Processing (ICIP).*

Dugelay, J.-L., & Petitcolas, F.A.P. (2000). Possible counter-attacks against random geometric distortions. *Proc. SPIE Security and Watermarking of Multimedia Contents II,* CA.

Hartung, F., & Girod, B. (1996). Digital watermarking of raw and compressed video. *Proc. SPIE digital compression technologies and systems for video commun., 2952,* 205–213.

Hartung, F., & Kutter, M. (1999). Multimedia watermarking technique. *Proc. IEEE, 8*(7), 1079–1106.

Hyvärinen, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks, 10,* 626–634.

Hyvärinen, A. (1999b). Survey on independent component analysis. *Neural Computing Surveys, 2,* 94–128.

Hyvärinen, A., & Oja, E. (1999). Independent component analysis: a tutorial. Retrieved from *www.cis.hut.fi/projects/ica/*

Inoue, H., Miyazaki, A., Yamamoto, A., & Katsura, T. (1998). A digital watermark based on the wavelet transform and its robustness on image compression. *Proc. Int. Conf. Image Processing (ICIP).*

Johnson, N.F., Duric, Z., & Jajodia, S. (1999). Recovery of watermarks from distorted images. *Preliminary Proc. of the Third Int. Information Hiding Workshop,* 361–375.

Kankanhalli, M.S., Rajmohan, & Ramakrishnan, K.R. (1998). Content based watermarking of images. *Proc. of the Sixth ACM International Multimedia Conference.*

Katzenbeisser, S., & Petitcolas, F.A.P. (2000). *Information hiding techniques for steganography and digital watermarking.* Boston: Artech House.

Koch, E., & Zhao, J. (1995). Towards robust and hidden image copyright labeling. *Proc. Workshop Nonlinear Signal and Image Processing.*

Kundur, D., & Hatzinakos, D. (1998). Digital watermarking using multiresolution wavelet decomposition. *Proc. of the Int. Conference on Acoustics, Speech, and Signal Processing, 5,* 2969–2972.

Kutter, M. (1998). Watermarking resisting to translation, rotation and scaling. *Proc. of SPIE Int. Symposium on Voice, Video, and Data Communications—Multimedia Systems and Applications, 3528,* 423–431.

Langelaar, C., Lubbe, J.C.A., & Lagendijk, R.L. (1997). Robust labeling methods for copy protection of images. *Proc. Electronic Imaging, 3022,* 298–309.

Lee, T.-W. (1998). *Independent component analysis: Theory and applications.* Kluwer Academic.

Liu, H., Kong, X.-W., Kong, X.-D., & Liu, Y. (2001). Content based color image adaptive watermarking scheme. *Proc. of IEEE International Symposium on Circuits and Systems, 2,* 41–44.

Lu, C.-S., & Mark, Liao H.-Y. (2001). Multipurpose watermarking for image authentication and protection. *IEEE Transaction on Image Processing, 10.*

Mintzer, F., & Braudaway, G. (1999). If one watermark is good, are more better? *Proc. of the International Conference on Acoustics, Speech, and Signal Processing, 4.*

Nikolaidis, N., & Pitas, I. (1998). Robust image watermarking in the spatial domain. *Signal Processing, 66,* 385–403.

Parhi, K.K., & Nishitani, T. (1999). *Digital signal processing for multimedia systems.* New York: Marcel Dekker.

Peter, P. (2001). *Digital image watermarking in the wavelet transform domain.* Unpublished master's thesis.

Piva, A., Barni, M., Bartoloni, E., & Cappellini, V. (1997). DCT-based watermark recovering without resorting to the uncorrupted original image. *Proc. IEEE Int. Conf. Image Processing (ICIP), 1.*

Podilchuk, C., & Zeng, W. (1997). Watermarking of the JPEG bitstream. *Proc. Int. Conf. Imaging Science, Systems, and Applications (CISST'97),* 253–260.

Ruanaidh, J.J.K.Ó, & Pun, T. (1997). Rotation, scale and translation invariant digital watermarking. *Proc. IEEE Int. Conf. Image Processing (ICIP'97), 1,* 536–539.

Ruanaidh, J.J.K.Ó, & Pun, T. (1998). Rotation, scale and translation invariant digital watermarking. *Signal Processing, 66*(3), 303–318.

Ruanaidh, J.J.K.Ó, Dowling, W.J., & Boland, F.M. (1996). Phase watermarking of digital images. *Proc. Int. Conf. Image Processing (ICIP'96), 3,* 239–242.

Schneier, B. (1995). *Applied cryptography* (2nd ed.). John Wiley and Sons.

Sheppard, N.P., Safavi-Naini, R., & Ogunbona, P. (2001). On multiple watermarking. *Proc. of ACM Multimedia 2001.*

Trappe, W., Wu, M., Wang, J., & Liu, K.J.R. (2003). Anti-collusion fingerprinting for multimedia. *IEEE Trans. on Signal Processing, 51*(4), 1069–1087.

Voloshynovskiy, S., Herrigel, A., Baumgaertner, N., & Pun, T. (1999). A stochastic approach to content adaptive digital image watermarking. *Proc. of Int. Workshop on Information Hiding.*

Wang, H.-J.M., Su, P.-C., & Kuo, C.-C.J. (1998). Wavelet-based digital image watermarking. *Optics Express, 3*(12), 491–496.

Xia, X., Boncelet, C., & Arce, G. (1997). A multiresolution watermark for digital images. *Proc. Int. Conf. Image Processing (ICIP'97), 1,* 548–551.

Xie, L., & Arce, G.R. (1998). Joint wavelet compression and authentication watermarking. *Proc. Int. Conf. Image Processing (ICIP'98).*

Yu, D., & Sattar, F. (2003). A new blind image watermarking technique based on independent component analysis. *Springer-Verlag Lecture Notes in Computer Science, 2613,* 51–63.

Yu, D., Sattar, F., & Ma, K.-K. (2002). Watermark detection and extraction using independent component analysis method. *EURASIP Journal on Applied Signal Processing—Special Issue on Nonlinear Signal and Image Processing (Part II).*

This work was previously published in Digital Watermarking for Digital Media, edited by J. Seitz, pp. 52-86, copyright 2005 by Information Science Publishing (an imprint of IGI Global).

Chapter 3.13
Digital Watermarking Schemes for Multimedia Authentication

Chang-Tsun Li
University of Warwick, UK

ABSTRACT

As the interconnected networks for instant transaction prevail and the power of digital multimedia processing tools for perfect duplication and manipulation increases, forgery and impersonation become major concerns of the information era. This chapter is intended to disseminate the concept of digital watermarking for multimedia authentication. Issues and challenges, such as security, resolution of tamper localization, and embedding distortion, of this technical area are explained first. Three main categories of watermarking approaches, namely fragile, semi-fragile, and reversible schemes, to the issues and challenges are then presented. Merits and limitations of specific schemes of each category are also reviewed and compared.

INTRODUCTION

As the interconnected networks for instant transaction prevail and the power of digital multimedia processing tools for perfect duplication and manipulation increases, forgery and impersonation become major concerns of the information era. As a result, the importance of authentication and content verification became more apparent and acute. In response to these challenges, approaches conveying the authentication data in digital media have been proposed in the last decade. Applications for multimedia authentication can be found in many areas. For example:

- **Medical image archiving**: The authentication data of patients can be embedded at the time when their medical images are taken by the hospital to protect the patients' rights when medical malpractice happens and has to be resolved in court.
- **Imaging/sound recording of criminal events**: Authentic imaging or recording of legally essential event or conversation could lead to breakthrough in criminal cases while maliciously tampered imaging/recording, if not detected, could result in wrong ruling.
- **Accident scene capturing for insurance and forensic purposes**: Similar applications of the technique as mentioned above could be

useful in protecting the rights of the parties including the insurance company involved in accidents or natural disasters.

- **Broadcasting**: During international crises, tampered or forged media could be used for propaganda and manipulating public opinion. Therefore, broadcasting is an area where multimedia authentication is applicable.
- **Military intelligence**: Multimedia authentication allows the military to authenticate whether the media they received do come from a legitimate source and to verify whether the content is original. Should the content be manipulated, an effective authentication scheme is expected to tell as much information about the manipulation as possible (e.g., location of tampering).

The aforementioned list is not intended to be exhaustive but just to identify some possible applications of multimedia authentication.

As Lou, Liu, and Li (2004) described, depending on the ways of conveying the authentication data for digital media, authentication techniques can be roughly divided into two categories: labeling-based techniques (Chen & Leiss, 1996; Friedman, 1993; Lin & Chang, 2001; Lou & Liu, 2000; Queluz, 2001; Schneider & Chang, 1996) and watermarking-based techniques (Hsieh, Li, & Wang, 2003; Li, Lou, & Chen, 2000; Li & Yang, 2003; Xie & Arce, 2001). The main difference between these two categories of techniques is that in labeling-based authentication, the authentication data or the signature of the medium is written into a separate file or a header that is separated from the raw data stored in the same file, while in watermarking-based authentication, the authentication data is embedded as watermark in the raw data itself.

Compared to watermarking-based techniques, labeling-based techniques potentially have the following advantages:

- The data-hiding capacity of labeling-based techniques is higher than that of watermarking.
- They can detect the change of every single bit of the image data if *strict* integrity has to be assured.

Given the above benefits, why would researchers propose watermarking approaches? The following are some of the issues regarding the use of labeling-based techniques:

- In labeling-based techniques, storing digital signatures in a separate file or in separate header segments of the file containing the raw data incurs significant maintenance overhead and may require extra transmission of the signature file.
- When the signed medium is manipulated, the embedded signature is not subjected to the same process of manipulation, which makes it difficult to infer what manipulation has been done and to pinpoint the temporal and spatial localities where tampering occurs.
- Traditional digital signatures used for labeling are not suitable for lossy or progressive transmission applications. For example, in light of network congestion in a progressive transmission scenario, low-priority layers of the medium (usually the high-frequency components or the details) are likely to be dropped, making the received data differ from the original. In this case, the received signature generated based on the original medium by the sender will not match its counterpart generated according to the received medium by the recipient. As a result, the received medium will fail authentication.
- Transcoding or converting the format of media-protected with labeling-based techniques is not always possible. For example, converting a JPEG image with its authentica-

tion data/signature stored in the header to an image format without any header segments means that the signature has to be discarded, making authentication impossible at a later stage.

On the contrary, watermarking-based approaches embed the authentication data into the raw data of the host media directly, which will be subjected to the same possible transformation the host media would undergo. Therefore, fragile and semi-fragile digital watermarking schemes do not have the first two aforementioned problems. Moreover, by sensibly designing the embedding algorithm, semi-fragile watermarking schemes can also circumvent the last two problems mentioned above. Nevertheless, readers are reminded that no superiority of the semi-fragile schemes over the fragile schemes is implied here. In deciding whether to make the scheme fragile or semi-fragile, the designer has to take the nature of applications and scenario into account since no single optimal scheme is available for all applications. Because of their merits, the rest of this chapter will focus on the design of watermarking-based schemes.

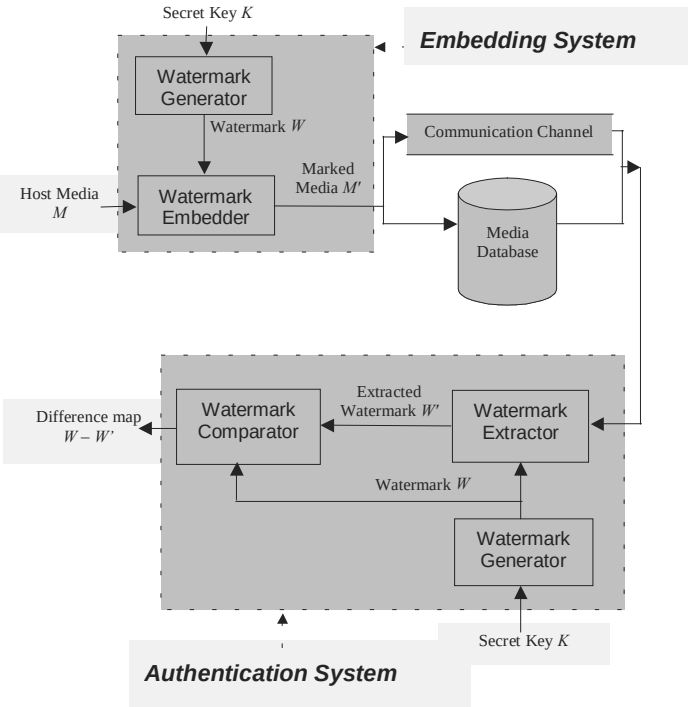
BACKGROUND

Various types of watermarking schemes have been proposed for different applications. For the purpose of copyright protection, embedded watermarks are expected to survive various kinds of manipulation to some extent, provided that the altered media is still valuable in terms of commercial significance or acceptable in terms of fidelity. Therefore, watermarking schemes for copyright protection are typically robust (Kutter, Voloshynovskiy, & Herrigel, 2000; Lu, Huang, Sze, & Liao, 2000; Trappe, Wu, & Liu, 2002; Wu & Liu, 2003), that is, they are trying to ignore or remain insensitive to the influence of malicious or unintentional attacks. On the other hand, in medical, forensic, broadcasting, and military applications where content verification and identity authentication are much more of a concern, more emphases are on the capability of the watermarking schemes to detect forgeries and impersonation. For example, the staff at a military headquarter always has to be sure that the digital images received come from the right people and that the contents are original. Therefore, this type of watermarks is usually fragile or semi-fragile and is expected to be sensitive to attacks (Li & Yang, 2003; van Leest, van der Veen & Bruekers, 2003; Winne, Knowles, Bull & Canagarajah, 2002; Wong & Memon, 2000; Xie & Arce, 2001). In addition to these two categories of schemes, some hybrid schemes incorporating a robust watermark and a fragile/semi-fragile watermark in attempt to provide copyright protection and authentication simultaneously have also been proposed (Deguillaume, Voloshynovskiy, & Pun, 2003).

This chapter deals with watermarking schemes for authentication purpose. A general authentication framework based on digital watermarking is illustrated in Figure 1. Usually, a secret key K available on both the embedding and authentication sides is used to generate a watermark to be embedded into the host media. The marked media is then delivered via the communication channel (e.g., Internet, satellite, etc.) or stored in a database. To authenticate the marked media, the same secret key is used to generate the original watermark so as to be used for extracting and comparing against the embedded version. The difference map, the output of the authentication system, tells the authenticity of the media.

Fragile/semi-fragile digital watermarking is about embedding a small amount of information, the watermark, into the host media so that the watermark can be extracted later from the marked media to authenticate the identity of the embedder and to verify the content integrity of the media. The watermark can be an ID number of the embedder, a visually meaningful logo, or a sequence of random numbers generated with a

Figure 1. A general authentication framework based on digital watermarking



secret key representing the embedder. The media can be text, audio, speech, image, video, multimedia, or a combination of them all.

When designing a watermarking scheme for digital media authentication, two categories of attacks have to be taken into consideration: content-targeting attacks and scheme-targeting attacks. Content-targeting attacks aim at manipulating the content of the media without taking into account the protection measures provided by the authentication algorithm. Common content-targeting attacks threatening digital media can be classified into two types:

- **Local tampering**: Typical examples of this type of tampering include removal of original objects/features and/or addition of new objects/features.
- **Global manipulations**: Typical examples of this type of manipulations include scaling, clipping, cropping, low-pass filtering, and histogram equalization.

Scheme-targeting attacks aim at defeating the security of the authentication scheme so that the scheme fails to detect malicious content manipulations. This type of attacks always goes along with content-targeting attacks because the ultimate goal of any malicious attacks is to tamper the content of the media. Typical examples of scheme-targeting attacks are as follows:

- Cover-up/cut-and-paste (Barreto, Kim, & Rijmen, 2002), which is an action of cutting one region/block of the image and pasting it somewhere else in the same or different image.
- Vector quantization attack (Wong & Memon, 2000) (also known as birthday attack [Barreto et al., 2002], the Holliman-Memon counterfeiting attack [Holliman & Memon, 2000], or collage attack [Fridrich, Goljan, & Memon, 2000]), which is devised on the basis of the so-called birthday paradox (Stallings, 1998, Appendix 8.A): *What is*

the minimum population size such that the probability that at least two of the people have the same birthday is greater than 0.5? According to birthday paradox, using a hash function that produces a bit string of length l, the probability of finding at least two image blocks that hash to the same output is greater than 0.5 whenever roughly $2^{l/2}$ watermarked image blocks are available. The idea of the attack is to forge a new watermarked image (a collage) from a number of authenticated images watermarked with the same key and the same logo/watermark by combining portions of different authenticated images while preserving their relative positions in the image. Fridrich, Goljan and Memon (2000) showed that counterfeiting is possible even when the logo is unknown to the attacker provided that a larger number of images watermarked with the same key are available.

- Transplantation attacks derived by Barreto, Kim, and Rijmen (2002) work as follows. Let $I'_A \rightarrow I'_B$ denote that the hashing of image block I'_B involves the information about I'_A. Now, if images I' and I'' have blocks with following dependence relationships:

$$\ldots \rightarrow I'_A \rightarrow I'_X \rightarrow I'_B \rightarrow I'_C \rightarrow \ldots$$
$$\ldots \rightarrow I''_A \rightarrow I''_X \rightarrow I''_B \rightarrow I''_C \rightarrow \ldots$$

and block I'_A is identical to I''_A, I'_B is identical to I''_B, and I'_C is identical to I''_C, but I'_X is not identical to I''_X. Then the block pairs (I'_X, I'_B) and (I''_X, I''_B) are interchangeable without being detected by schemes adopting deterministic dependence (Li et al., 2000; Wong & Memon, 2000; Yeung & Minzter, 1997), that is, the information involved or dependent upon is deterministic. Barreto et al. (2002) further indicated that merely increasing the number of dependencies could not thwart the transplantation attack. For example, let $I_A \leftrightarrow I_B$ denote that the hashing of each block involves the information about the other. Now, if the following dependence relationships exist

$$\cdots \leftrightarrow I'_A \leftrightarrow I'_B \leftrightarrow I'_X \leftrightarrow I'_C \leftrightarrow I'_D \leftrightarrow \cdots$$
$$\cdots \leftrightarrow I''_A \leftrightarrow I''_B \leftrightarrow I''_X \leftrightarrow I''_C \leftrightarrow I''_D \leftrightarrow \cdots,$$

the triplet (I'_B ,I'_X , I'_C) and (I''_B ,I''_X , I''_C) are interchangeable if block I'_D is also identical to I''_D.

Based on the above discussions, it is clear that content-targeting attacks are naïve and can only succeed when the attacked media is *not* watermarked. Scheme-targeting attacks are more sophisticated and intended for making content-targeting attacks undetectable.

To be considered effective, a fragile/semi-fragile watermarking scheme must have the essential capabilities of detecting content-targeting attacks. Moreover, an effective fragile/semi-fragile watermarking scheme should show no security gaps to various kinds of scheme-targeting attacks such as cover-up, transplantation, and vector quantization attacks. Block-wise dependence is recognized as a key requirement to thwart vector quantization (Fridrich et al., 2000; Holliman & Memon, 2000; Li et al., 2000; Li & Yang, 2003; Wong & Memon, 2000). However, it is also proved that the dependency with deterministic context is susceptible to transplantation attack or even simple cover-up attack (Barreto et al., 2002). Nevertheless, Li and Yang (2003) pointed out that even nondeterministic block-wise dependency (i.e., the neighboring relationship among individual pixels in the block is nondeterministic) as adopted in Fridrich, Goljan, and Baldoza (2000) is still vulnerable to cropping attack.

Although spatial-domain approaches are effective for the applications where lossy compression is not acceptable (e.g., medical imaging), lossy compression standards such as JPEG and MPEG are commonly adopted for Internet transmission and multimedia storage in order to make efficient

use of bandwidth and disk space. These two requirements make transform-domain approaches desirable. It is common in transform-domain schemes to watermark only a few selected transform coefficients in order to minimize embedding distortion. However, Hsieh, Li, and Wang (2003) pointed out that leaving most of the coefficients unmarked results in a wide-open security gap for attacks to be mounted on them. A solution Hsieh et al. (2003) proposed is to implicitly watermark all the coefficients by registering/blending the zero-valued coefficients in the watermark and involving the unembeddable coefficients during the process of embedding the embeddable ones.

However, fragile watermarking is intolerant not only to malicious attacks but also to content-preserving or incidental operations (e.g., further compression, transcoding, bit rate scaling, and frame rate conversion), which do not change the semantics or content of the media. Those content-preserving operations are sometimes necessary in many multimedia applications wherein fragile watermarking is not practical. Semi-fragile watermarking (Kundur & Hatzinakos, 1999; Xie & Arce, 2001) is the technique that allows those content-preserving operations while intolerant to malicious content-altering operations such as removal of original objects from the media.

In some applications, such as medical imaging and forensic image archiving, even imperceptible distortion due to watermark embedding is not acceptable. For example, in legal cases of medical malpractice, any distortion of an image, even if it is a result of watermark embedding itself, would cause serious debate in court. Therefore, the capability of recovering the original unwatermarked media from the authenticated watermarked version is of significant value for this type of application. A watermarking scheme with this capability is referred to as reversible (van Leest, van der Veen, & Bruekers, 2003), erasable (Cox, Miller, & Jeffrey, 2002). or invertible watermarking (Fridrich, Goljan, & Du, 2001).

In addition, it is more practical for a watermarking scheme to be able to verify the authenticity and integrity of the media without referring to the original versions. This characteristic is commonly called obliviousness or blind verification. In a more restrictive sense, obliviousness can also mean that no a priori knowledge (e.g., image index) about the image is required in the authentication process.

Low false-positive and false-negative rates are also important factors for effective schemes. False-positive rate is the occurrence rate that the watermark detector extracts a watermark that was *not* actually embedded. On the contrary, false-negative rate is the occurrence rate that the watermark detector fails to extract an embedded watermark. Low false-positive and false-negative rates are usually in conflict with low embedding distortion because reducing false-positive and false-negative rates usually means increasing the amount of watermark, which inevitably will inflict higher distortion on the quality of the watermarked media.

The aforementioned challenges and attacks do not constitute an exhaustive list since more new attacks are expected to be devised in the future. Nevertheless, at present, an effective watermarking scheme for authentication purpose should have the capability of thwarting known attacks.

WATERMARKING APPROACHES TO AUTHENTICATION

This section is intended to introduce some solutions to the problems posed previously. Several watermarking approaches will be discussed.

Fragile Watermarking Schemes

Among the proposed spatial-domain fragile watermarking techniques, the Yeung-Mintzer scheme (Yeung & Minzter, 1997) is one of the earliest and frequently cited. In Yeung and Minzter (1997),

the watermark is a visually significant binary logo, which is much smaller than the image to be marked and is used to form a binary image as big as the image. Watermark embedding is conducted by scanning each pixel and performing the watermark extraction function based on a lookup table generated with a secret key. If the extracted watermark bit is equal to the authentic watermark bit, the pixel is left unchanged; otherwise, the gray scale of the pixel is adjusted until the extracted watermark bit is equal to the authentic one. Because of its pixel-wise scanning fashion, local tampering can be localized to pixel accuracy. The pixel-wise watermarking fashion is actually a special case of the block-wise style with block size equal to 1.

However, due to the lack of interrelationship among neighboring pixels during the watermarking process, their scheme is vulnerable to cover-up attacks when there are local features surrounded by a relatively larger smooth background. For example, without knowing the secret key, a handgun on the floor of a criminal scene can still be covered up by pasting a patch taken from the background. The scheme is also vulnerable to vector quantization attack (Fridrich et al., 2000; Holliman & Memon, 2000; Wong & Memon, 2000). Another attack derived by Fridrich et al. (2000) that can be mounted against the Yeung-Mintzer scheme is that the lookup table and the binary logo can be inferred when the same lookup table and logo are reused for multiple images.

Another well-known fragile watermarking technique is Wong's public-key scheme reported in Wong (1998). In this scheme, the gray scales of the least significant bits (LSBs) of the original image are first set to zero. Then the LSB-zeroed image is divided into blocks of the same size as that of a watermark block. The image size together with each LSB-zeroed image block is then provided as inputs to a hash function and the output together with the watermark block are subjected to an exclusive-or (XOR) operation. The result of the XOR operation is then encrypted using a private key of the RSA public-key cryptosystem and embedded in the LSBs of the original image. To verify the integrity of the received image, it is again divided into blocks and the encrypted data embedded in the LSBs of each block is then extracted and decrypted with a public key. Meanwhile, the LSB-zeroed version of each received image block together with the image size of the transmitted image are again taken as inputs to the same hash function as that used by the image sender. The output of the hash function and decrypted data are subjected to the XOR operation in order to reveal the watermark. The revealed watermark is then compared with the perfect watermark in the preestablished database at the receiver side. This scheme marries cryptography and watermarking elegantly and indeed works well in detecting cropping and scaling.

However, like the Yeung-Mintzer scheme, this method is also block-wise independent, and, therefore, vulnerable to cover-up and vector quantization attacks. Since the block size of Wong's scheme is 64, according to birthday paradox, given 2^{32} blocks, vector quantization attack can be successful with relatively high probability. This is possible in the applications of medical image archiving where large image database is maintained. Due to the lack of mutual dependence among neighboring blocks during the watermarking process, this scheme is also vulnerable to transplantation attacks. Moreover, the output length of the hash function sets the lower bound on the block size. Thus, the tampering localization accuracy is limited.

To thwart vector quantization attack, Wong and Memon (2000) proposed an improved scheme by adding an image index and a block index to the inputs of the hash function. With this new version, to forge each block, the choices for the attacker are now limited to only the blocks from all authenticated images with the same block index. Adding the image index is one step further to secure the scheme against vector quantization attack. In this case, the image index is just like

a unique serial number of the image; therefore, vector quantization cannot succeed. However, this idea works at the expense of requiring the verifier to have the a priori knowledge about the image index, which limits its applicability to some extent. For example, an intelligence agent in a hostile territory has to send the index of the image he/she wants to transmit through a secure channel to the verifier.

Recognizing the importance of establishing dependence among neighboring pixels or blocks, Li, Lou, and Chen (2000) proposed a scheme that uses a binary feature map extracted from the underlying image as watermark. The watermark is then divided into blocks of size 32×16 pixels. Block-wise dependence is established by blending the neighboring blocks before encrypting and embedding into LSBs of the image. This method is effectively resistant to vector quantization and cover-up attacks and requires no a priori knowledge of the original image. However, the accuracy of localization is limited by the block size. Moreover, like Wong's (1998) scheme, this scheme is also vulnerable to transplantation attacks because the contextual dependence is established based on deterministic information. To circumvent these drawbacks, Li and Yang (2003) further proposed a scheme that is immune to transplantation attacks and is significantly accurate in locating tampering. To watermark the underlying image, the scheme adjusts the gray scale of each pixel by an imperceptible quantity according to the consistency between a key-dependent binary watermark bit and the parity of a bit stream converted from the gray scales of a secret neighborhood. The members of the secret neighborhood are selected according to the watermark sequence generated with the secret key, and therefore cannot be deterministically reestablished by the attacker. However, it is a spatial-domain approach, which is not suitable for transform-domain applications.

Although there are some transform-domain schemes reported in the literature, a common security gap inherent in many of them (Winne, Knowles, Bull, & Canagarajah, 2002; Wu & Liu, 1998; Xie & Arce, 2001) is that they neither explicitly nor implicitly watermark all the transform coefficients. As a result, manipulation of those unwatermarked coefficients will go unnoticed. For example, in the wavelet transform-domain approach proposed by Winne, Knowles, Bull, and Canagarajah (2002), to minimize embedding distortion and maintain high localization accuracy, only the coefficients of the high frequency subbands at the finest scale of the luminance component are watermarked. All the other coefficients and components are neither watermarked nor involved during the watermarking process of the embeddable coefficients. In Xie and Arce (2001), to make the scheme semi-fragile, only the LL component of the coarsest scale (i.e., the approximate of the original image) is involved in generating the signature, which is then used as the watermark. To minimize embedding distortion, only the coefficients of the finest scale are watermarked. Consequently, tampering the coefficients in other subbands and scales will certainly go undetected. For example, locally tampering the three unwatermarked high-frequency subbands at the coarsest scale that are not involved in generating the signature is highly likely to change or at least destroy the semantic meaning of the watermarked image without raising alarm.

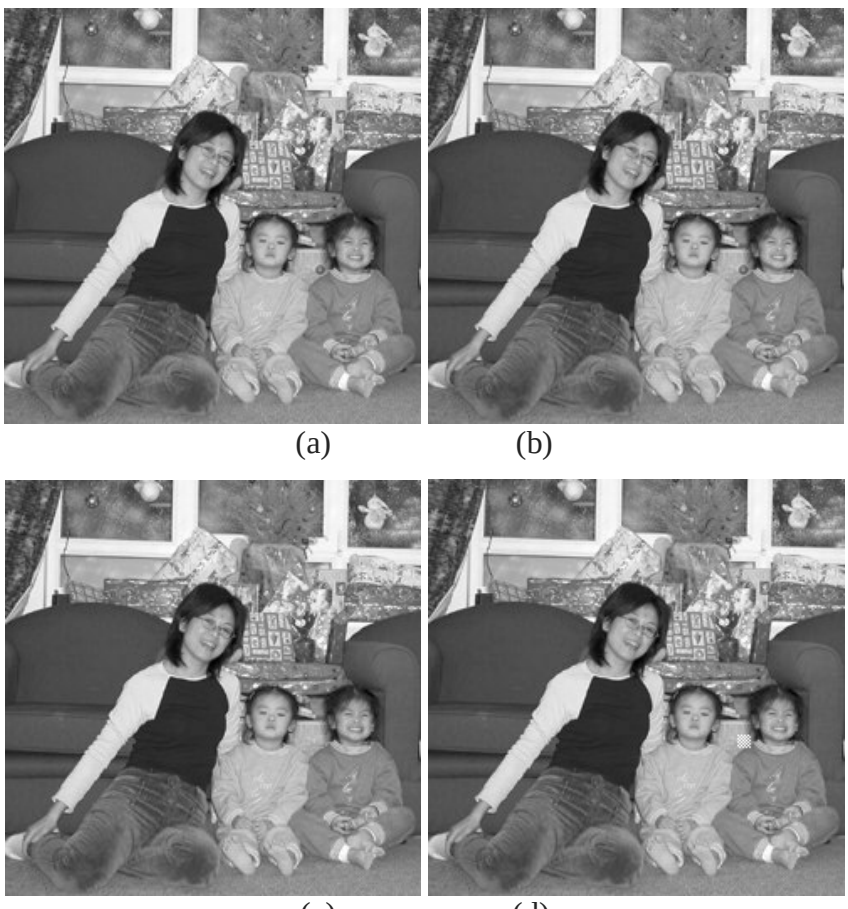
Given the limitations of the reviewed schemes, Hsieh, Li, and Wang (2003) designed a transform-domain scheme, which is immune to the aforementioned attacks and provides protection for *all* the transform coefficients without explicitly watermarking all of them. To embed the watermark, the target image X is first DCT transformed and quantized. A binary image A as big as X is generated with a secret key. A second binary image, B, is then created so that all of its pixels corresponding to the nonzero-valued coefficients are set to 1 and the others set to 0. B is intended to serve the purpose of registering the positions of the zero-valued coefficients. A binary watermark W is then created by taking the

result of XOR operation on the binary images *A* and *B*. Like *X*, *W* is also divided into blocks of 8×8 pixels. For each DCT block X_i, four nonzero coefficients $X_i(h)$, $X_i(h\text{-}1)$, $X_i(h\text{-}2)$, and $X_i(h\text{-}3)$ with their frequencies lower than or equal to a middle frequency *h* are identified as watermarkable. The four selected coefficients $X_i(j)$, $j \in [h\text{-}3, h]$, are modulated based on their corresponding watermark bits $W_i(j)$ and a secret sum $S_i(j)$ such that Equation (1) is satisfied.

$$\text{Parity}\,(S_i(j) \copyright X_i(j)) = W_i(j) \qquad (1)$$

where *Parity*(×) is a function which returns 1 or 0 as output to indicate that the number of '1' bits of its argument is *odd* or *even*. © is an operator that concatenates $S_i(j)$ and $X_i(j)$ expressed in two's complement format. The secret sum $S_i(j)$ is the sum of the nonzero unwatermarkable coefficients selected according to their corresponding watermark bits and $W_i(j)$ from a neighborhood $N_i(j)$ consisting of the DCT block X_i and its eight closest neighboring blocks. It canbe expressed as

Figure 2. (a) Original image. (b) Watermarked image. (c) Attacked image. The doorknob between the twins' heads in the watermarked image has been masked. (d) Authentication result. The shaded block between the twins' heads indicates that the received image has been locally tampered with.



(a) (b)

(c) (d)

$$S_i(j) = \sum_{m \in N_i(j)} \sum_{n \in [h-3,h]} (W_m(n) \oplus W_i(j)) \cdot X_m(n) \quad (2)$$

The watermarking process repeats until all the blocks are marked. To authenticate and verify the received image, the verifier performs the same operations as applied on the embedding side in the reversed order to extract the embedded watermark and compares it with the original watermark generated in the same manner as that adopted by the embedder. A set of experimental results of this scheme is shown in Figure 2. Figure 2(a) is the original image while Figure 2(b) is the watermarked image. These two images show that the embedding distortion is invisible to human visual system (HVS). Figure 2(c) shows that the doorknob between the twins' heads has been removed (compare the difference between Figure 2(b) and 2(c)). This is a typical example of localized attack. Detected by the authentication scheme, the shaded block between the twins' head indicates that the image has been locally tampered with. These two figures show that the scheme is capable of localizing tampering to high accuracy.

Semi-Fragile Watermarking Schemes

One characteristic of the aforementioned fragile watermarking scheme is its zero tolerance to any types of changes no matter how small they are. This characteristic makes the fragile scheme unsuitable in many applications where content-preserving manipulation is necessary. In order to make efficient use of bandwidth and storage, media is often transmitted or stored in compressed formats according to some specific standards such as JPEG and MPEG. Transcoding is also a common practice to convert media from one format to another (e.g., from JPEG to TIFF). Compression and transcoding are deemed acceptable in many Internet and multimedia applications as they preserve the content. However, fragile watermarking schemes do not differentiate content-preserving operations from malicious tampering. Therefore, to meet the needs for authenticating the compressed or transcoded watermarked media, it is desirable to have semi-fragile schemes that are sensitive to malicious manipulations while tolerant to content-preserving operations.

Kundur and Hatzinakos (1999) developed a semi-fragile watermarking scheme for image authentication, which embeds the watermark by quantizing the wavelet coefficients to a predetermined degree. The scheme works as follows. They defined a quantization function $Q(f)$ as

$$Q(f) = \begin{cases} 0, & \text{if } \left\lfloor \frac{f}{\delta 2^l} \right\rfloor \text{ is even} \\ 1, & \text{if } \left\lfloor \frac{f}{\delta 2^l} \right\rfloor \text{ is odd} \end{cases} \quad (3)$$

where f stands for any coefficient, $\lfloor \cdot \rfloor$ is the floor function which returns the largest integer smaller than or equal to its argument, l is the index of the decomposition level, and d, the quantization step, is a positive integer. To watermark the image, the L-level Haar wavelet transform is performed first. Then for all the coefficients except the ones of the approximation of the image (i.e., the coefficients of the lowest-frequency subband) are subjected to the selection by using a secret key *ckey*. For each selected coefficient $f(i)$, if Equation 4 does not hold, the coefficient is adjusted according to Equation 5. Equations 4 and 5 are deined as follows:

$$Q(f(i)) = w(i) \oplus qkey(i) \quad (4)$$

where $w(i)$ is the ith bit of the watermark sequence, $qkey(i)$ is a function of the local component of the image around pixel i which returns either value 0 or 1 and is intended for increasing the security of the scheme, and $\oplus$ is the XOR operator,

$$f(i) := \begin{cases} f(i) - \delta 2^l, & \text{if } f(i) > 0 \\ f(i) + \delta 2^l, & \text{if } f(i) \le 0 \end{cases} \quad (5)$$

where the operator := stands for assignment. After the watermarking process as described above is finished, inverse Haar wavelet transform is performed on the watermarked coefficients to create the watermarked image.

As most of the existing digital image formats require that the gray level of the pixels must be an integer; when the inverse wavelet transform is applied to the watermarked coefficients, the resulting gray level of the watermarked image pixel must be rounded to integer values. However, the rounding operation may result in changing the watermark because of this tiny numerical modification. To solve this problem, Kundur and Hatzinakos (1999) chose Haar wavelet transform, exploiting the property that the coefficients at each decomposition level l are rational numbers of the form $r/2^l$ where r is an integer value. Watermarking the coefficients by using Equation 3 and adjusting the coefficients by a multiple of 2^l according to Equation 5, the gray levels of the inverse wavelet transform are guaranteed to be integers. We can also see from Equations 3 and 5 that the quantization step δ determines the degree of distortion and sensitivity of the scheme to changes in the image. A smaller value of d inflicts less significant distortion on the visual quality of the image while making the scheme less tolerant to changes.

Depending on the applications, the watermarked image may be subjected to some kind of content-preserving operations (e.g., lossy compression) before the image is transmitted through the communication channel or stored in the database. Therefore, to verify its authenticity, the received or retrieved watermarked image has to be transformed or decoded back to its spatial domain wherein the watermark extraction described as follows can take place. To extract the embedded watermark, the L-level Haar wavelet transform exactly as carried out in the embedding process is performed first. Then for all the coefficients except the ones of the approximation of the image are subjected to the selection by using a secret key *ckey*. For each selected coefficient $f(i)$, the corresponding watermark bit $\tilde{w}(i)$ is extracted according to

$$\tilde{w}(i) = Q(f(i)) \oplus qkey(i) \qquad (6)$$

A tamper assessment function (TAF) is then calculated according to

$$\mathrm{TAF}(\tilde{w}, w) = \frac{1}{N_w}\sum_{i=1}^{N_w} w(i) \oplus \tilde{w}(i) \qquad (7)$$

where w and $\tilde{w}$ are the original and extracted watermark sequences, respectively, and N_w is the length of the watermark sequences. The received/retrieved image is deemed authentic if the value of $\mathrm{TAF}(\tilde{w}, w) < T$, where $0 \le T \le 1$ is a user-defined threshold. Otherwise, the changes to the image are considered content preserving and acceptable. The value of T is application dependent. We can see that the higher its value is, the more sensitive the scheme becomes. Experiments conducted by Kundur and Hatzinakos (1999) suggest that a value of approximately 0.15 for T allows the scheme to be robust against high-quality JPEG compression and be able to detect additional tampering.

Although desirable, it is difficult to draw a clear boundary between acceptable and malicious manipulations. The designer has to bear in mind what the application is so that he/she can differentiate acceptable distortions from malicious ones.

Reversible Watermarking Schemes

One limitation of watermarking-based authentication schemes is the distortion inflicted on the host media by the embedding process. Although the distortion is often insignificant, it may not be acceptable for some applications, especially in the areas of medical imaging. Therefore, watermarking scheme capable of removing the distortion and recovering the original media after passing the authentication is desirable. Schemes with this capability are often referred to as reversible

watermarking schemes (also known as invertible [Fridrich et al., 2001] or erasable watermarking [Cox et al., 2002]). None of the algorithms mentioned previously are reversible. Usually, a reversible scheme performs some type of lossless compression operation on the host media in order to make space for hiding the compressed data and the Message Authentication Code (MAC) (e.g., hash, signature, or some other feature derived from the media) used as the watermark. To authenticate the received media, the hidden information is extracted and the compressed data is decompressed to reveal the possible original media. MAC is then derived from the possible original media. If the newly derived MAC matches the extracted one, the possible original media is deemed authentic/original. Two interesting reversible schemes are introduced as follows.

For the scheme proposed by Fridrich et al. (2001), first, the 128-bit hash of all the DCT coefficients is calculated. A number of middle-frequency coefficients are then selected from each DCT block. The least significant bits of the selected coefficients are losselessly compressed when the coefficients are scanned in a secretly determined order. The lossless compression stops when enough space has been created for embedding the hash. The compressed bit stream and the hash are then concatenated and replace the LSBs of the selected coefficients. To verify the authenticity, the verifier follows the same protocol to select the same middle-frequency coefficients in order to extract the compressed bit stream and hidden hash H from their LSBs. The extracted compressed bit stream is then decompressed and used to replace LSBs of those selected middle-frequency coefficients. The same hash function is applied to all the coefficients to obtain H'. If H' equals H, the received image is deemed authentic and the LSBs of the received image are replaced with the decompressed bit stream to yield the original. Despite its simplicity, the hash output conveys only global information about the image, that is, the signature of the image, with no local information. When a local attack is launched against the coefficients, their algorithm can only tell that the image is not authentic without being able to locate the position where the tampering occurs.

Van Leest, van der Veen, and Bruekers (2003) proposed another reversible watermarking scheme based on a transformation function that introduces "gaps" in the image histogram of image blocks. The transformation function maps the gray level of the input pixel to a unique output value so that one or two values in the range are not used, thus leaving one or two "gaps." The gaps are then used to hide the watermark. For example, a possible transformation function is one that maps the domain of [0, 1, 2, …, x, x+1, x+2, x+3, …, x'] to the range of [0, 1, 2, …, x, x+2, x+3, x+4, …, x' + 1], leaving value x+1 unmapped in the range. The scheme then embeds a "1" watermark bit by increasing the gray level of any pixel with a gray level of x by 1 to make it equal to x +1 and a "0" by not changing anything. We can see that after the embedding process is done, the gaps corresponding to gray level x+1 is partially filled and the embedding capacity is determined by the occurrences of gray level x. By allowing more gaps, higher embedding capacity can be gained at the expense of greater distortion to the visual quality. Along with some overhead information indicating the whereabouts of the gaps, the watermark verifier can extract the information and restore the original image in a bit-exact manner. Experiments demonstrated that embedding rates of approximately 0.06–0.60 bits per pixel could be achieved at PSNR levels of 45–50 dB. One drawback of this scheme is its need for the overhead information and the protocol to be hidden in the image. Moreover, a potential security loophole in the scheme is that given the fact that the computational cost for extracting the watermark is insignificant; an attacker can defeat the scheme by exhausting all the 256 possible gray level assuming that the gray level being tried is the gap.

CONCLUSION

This chapter is about the use of digital watermarking for multimedia authentication. The first section discussed the pressing needs for authenticating digital media in this information era and the two main categories of authentication techniques employed to meet these needs, namely labeling-based techniques and watermarking-based techniques. Characteristics of these two categories of techniques were compared and the reasons why watermarking is preferred in some applications were presented.

The second section identified some common attacks and classified them into content-targeting attacks and scheme-targeting attacks. How the attacks could be mounted on the media and what requirements have to be met in order to thwart those attacks were also explained.

In the third section, depending on the properties of the watermarking schemes and the desirable requirements of applications, digital watermarking schemes were broadly classified into three categories, namely fragile, semi-fragile, and reversible. Some existing schemes of each category were described in detail.

Based on the discussions made in the previous sections, it is observed that no single universal solution to all problems currently exist and is unlikely to be found in the future. The solutions are more likely to remain application dependent and the trade-offs between the conflicting requirements of low distortion, low false-positive and negative rates, and robustness to acceptable manipulations still have to be made. The authors expect that the future trends in this field are increasing the localization accuracy, identifying the type of tampering, and restoring the original media.

REFERENCES

Barreto, P.S.L.M., Kim, H.Y., & Rijmen, V. (2002). Toward secure public-key blockwise fragile authentication watermarking. *IEE Proceedings—Vision, Image and Signal Processing, 148*(2), 57–62.

Chen, F., & Leiss, E.L. (1996). Authentication for multimedia documents. *Proceedings of Conferencia Latinoamérica de Informática*, 613–624.

Cox, I., Miller, M., & Jeffrey, B. (2002). *Digital watermarking: Principles and practice*. Morgan Kaufmann.

Deguillaume, F., Voloshynovskiy, S., & Pun, T. (2003). Secure hybrid robust watermarking resistant against tampering and copy-attack. *Signal Processing, 83*(10), 2133–2170.

Fridrich, J., Goljan, M., & Baldoza, A.C. (2000). New fragile authentication watermark for images. *Proceeding of the IEEE International Conference on Image Processing, I,* 446–449.

Fridrich, J., Goljan, M., & Du, R. (2001). Invertible authentication watermark for JPEG images. *Proceeding of the IEEE International Conference on Information Technology,* 223–227.

Fridrich, J., Goljan, M., & Memon, N. (2000). Further attack on Yeung-Mintzer watermarking scheme. *Proceeding of the SPIE Conference on Security and Watermarking of Multimedia Content, II,* 428–437.

Friedman, G.L. (1993). The trustworthy digital camera: Restoring credibility to the photographic image. *IEEE Transactions on Consumer Electronics, 39*(4), 905–910.

Holliman, M., & Memon, N. (2000). Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. *IEEE Transactions on Image Processing, 9*(3), 432–441.

Hsieh, T.-H., Li, C.-T., & Wang, S. (2003). Watermarking scheme for authentication of compressed image. *Proceeding of the SPIE International Conference on Multimedia Systems and Applications, VI,* 1–9.

Kundur, D., & Hatzinakos, D. (1999). Digital watermarking for telltale tamper proofing and authentication. *Proceedings of IEEE, 87*(7), 1167–1180.

Kutter, , Voloshynovskiy, S., & Herrigel, A. (2000). Watermark copy attack. *Proceeding of the SPIE International Conference on Security and Watermarking of Multimedia Content, II*, 371–380.

Li, C.-T., & Yang, F.-M. (2003). One-dimensional neighbourhood forming strategy for fragile watermarking. *Journal of Electronic Imaging,12*(2), 284–291.

Lin, C.-Y., & Chang, S.-F. (2001). A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Transactions on Circuits and Systems of Video Technology, 11*(2), 153–168.

Li, C.-T., Lou, D.-C., & Chen, T.-H. (2000). Image authentication via content-based watermarks and a public key cryptosystem. *Proceedings of the IEEE International Conference on Image Processing, III*, 694–697.

Lou, D.-C., & Liu, J.-L. (2000). Fault resilient and compression tolerant digital signature for image authentication. *IEEE Transactions on Consumer Electronics, 46*(1), 31–39.

Lou, D.-C., Liu, J.-L., & Li, C.-T. (2004). Digital signature-based image authentication. In C.S. Lu (Ed.), *Multimedia security: Steganography and digital watermarking techniques for protection of intellectual property*. Hershey, PA: Idea Group.

Lu, C.S., Huang, S.K., Sze, C.J., & Liao, H.Y. (2000). Cocktail watermarking for digital image protection. *IEEE Transactions on Multimedia, 2*(4), 209–224.

Queluz, M.P. (2001). Authentication of digital images and video: Generic models and a new contribution. *Signal Processing: Image Communication, 16*, 461–475.

Schneider, M., & Chang, S.-F. (1996). Robust content based digital signature for image authentication. *Proceedings of the IEEE International Conference on Image Processing, III*, 227–230.

Stallings, W. (1998). *Cryptography and network security: Principles and practice*. Prentice Hall.

Trappe, W., Wu, M., & Liu, K.J. (2002). Collusion-resistant fingerprinting for multimedia. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3309–3312.

van Leest, A., van der Veen, M., & Bruekers, F. (2003). Reversible image watermarking. *Proceedings of the IEEE International Conference on Image Processing, II*, 731–734.

Winne, D.A., Knowles, H.D., Bull, D.R., & Canagarajah, C.N. (2002). Digital watermarking in wavelet domain with predistortion for authenticity verification and localization. *Proceeding of the SPIE Conference on Security and Watermarking of Multimedia Contents, IV*, 349–356.

Wong, P.-W. (1998). A public key watermark for image verification and authentication. *Proceeding of the IEEE International Conference on Image Processing, I*, 455–459.

Wong, P.-W., & Memon, N. (2000). Secret and public key authentication watermarking schemes that resist vector quantization attack. *Proceeding of the SPIE Conference on Security and Watermarking of Multimedia Contents, II*, 417–427.

Wu, M., & Liu, B. (1998). Watermarking for image authentication. *Proceeding of the IEEE International Conference on Image Processing, II*, 437–441.

Wu, M., & Liu, B. (2003). Data hiding in image and video: 1. Fundamental issues and solutions. *IEEE Transactions on Image Processing, 12*(6), 685–695.

Xie, L., & Arce, G.R. (2001). A class of authentication digital watermarks for secure multimedia communication. *IEEE Transactions on Image Processing, 10*(11), 1754 –1764.

Yeung, M., & Minzter, F. (1997). An invisible watermarking technique for image verification. *Proceeding on the IEEE International Conference on Image Processing, I,* 680–683.

This work was previously published in Digital Watermarking for Digital Media, edited by J. Seitz, pp. 30-51, copyright 2005 by Information Science Publishing (an imprint of IGI Global).

Section 4

Utilization and Application

This section introduces and discusses a variety of the existing applications of multimedia technologies that have influenced education, science, and even music and proposes new ways in which multimedia technologies can be implemented within organizations and in society as a whole. Within these selections, particular multimedia applications, such as face recognition technology and educational, are explored and debated. Contributions included in this section provide excellent coverage of today's multimedia environment and insight into how multimedia technologies impact the fabric of our present-day global village.

Chapter 4.1
Integrated-Services Architecture for Internet Multimedia Applications

Zhonghua Yang
Nanyang Technological University, Singapore

Yanyan Yang
University of California, USA

Yaolin Gu
Southern Yangtze University, China

Robert Gay
Nanyang Technological University, Singapore

A HISTORICAL PERSPECTIVE

The Internet has gone from near invisibility to near ubiquity and penetrated into every aspect of society in the past few years (Department of Commerce, 1998). The application scenarios have also changed dramatically and now demand a more sophisticated service model from the network. In the early 1990s, there was a large-scale experiment in sending digitized voice and video across the Internet through a packet-switched infrastructure (Braden, Clark, & Shenker, 1994). These highly visible experiments have depended upon three enabling technologies: (a) Many modern workstations now come equipped with built-in multimedia hardware, (b) IP multicasting, which was not yet generally available in commercial routers, is available, and (c) highly sophisticated digital audio and video applications have been developed. It became clear from these experiments that an important technical element of the Internet is still missing: Real-time applications often do not work well across the Internet. The Internet, as originally conceived, offers only a very simple quality-of-service (QoS), point-to-point, best-effort data delivery. However, for a real-time application, there are two aspects of the problem with using this service model. If the

sender and/or receiver are humans, they simply cannot tolerate arbitrary delays; on the other hand, if the rate at which video and audio arrive is too low, the signal becomes incomprehensible. To support real-time Internet applications, the service model must address those services that relate most directly to the time of delivery of data. Real-time applications like video and audio conferencing typically require stricter guarantees on throughput and delay. The essence of real-time service is the requirement for some service guarantees in terms of timing. In response to these demands of real-time multimedia applications, the Internet Engineering Task Force (IETF) has significantly augmented the Internet protocol stack based on the Internet integrated-services model, which is the focus of this article.

THE INTERNET INTEGRATED-SERVICES MODEL

An Internet *service model* consists of a set of service commitments; that is, in response to a service request, the network commits to deliver some service. The Internet is conventionally designed to offer a very simple service model, best effort, providing no guarantee on the correct and timely delivery of data packets. Each request to send is honored by the network *as best as it can*. This is the worst possible service: Packets are forwarded by routers solely on the basis that there is a known route, irrespective of traffic conditions along that route. This simplicity has probably been one of the main reasons for the success of IP technology. The best-effort service model, combined with an efficient transport-layer protocol (TCP [transmission-control protocol]), is perfectly suited for a large class of applications, which tolerate variable delivery rates and delays. This class of applications is called elastic applications.

However, demanding real-time applications require more sophisticated service models beyond the best effort. There has been a great deal of effort since 1990 by IETF to add a broad range of services to the Internet service model, resulting in the Internet integrated service model (Braden et al., 1994; Crowcroft, Handley, & Wakeman, 1999). The Internet integrated services model defines five classes of services that should satisfy the requirements of the vast majority of future applications.

1. *Best effort*: As described above, this is the traditional service model of the Internet.
2. *Fair*: This is an enhancement of the traditional model where there are no extra requests from the users, but the routers attempt to partition network resources in some fair manner. This is typically implemented by adopting a random-drop policy when encountering overload, possibly combined with some simple round-robin serving of different sources.
3. *Controlled load*: This is an attempt to provide a degree of service guarantee so that a network appears to the user as if there is little other traffic, and it makes no other guarantees. The admission control is usually imposed so that the performance perceived is as if the network were overengineered for those that are admitted.
4. *Predictive service*: This service gives a delay bound that is as low as possible, and at the same time, is stable enough that the receiver can estimate it.
5. *Guaranteed service*: This is where the delay perceived by a particular source or to a group is bounded within some absolute limit. This service model implies that resource reservation and admission control are key building blocks of the service.

The level of QoS provided by these enhanced QoS classes is programmable on a per-flow basis, and end-to-end QoS commitment for the data flow is built using a unified multimedia protocol stack and through resource reservation.

THE INTERNET MULTIMEDIA PROTOCOL ARCHITECTURE

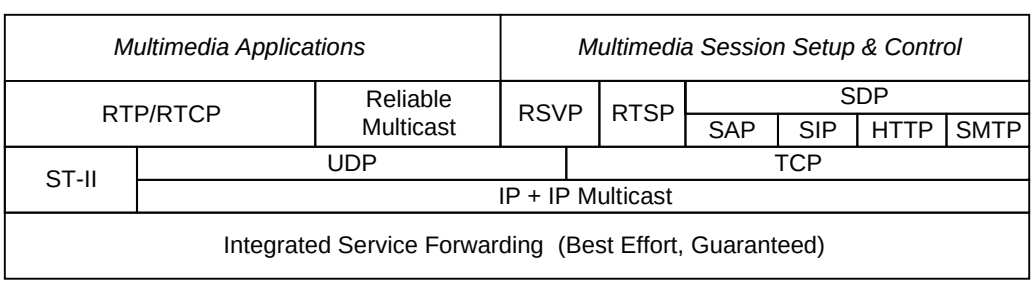
The integrated-services Internet offers a class of service models beyond the TCP/IP's (Internet protocol) best-effort service, and thus it imposes strict new requirements for a new generation of Internet protocols. The set of Internet real-time protocols, which constitute the Internet multimedia protocol architecture, represents a new style of protocols. The new style of protocols follows the proposed principles of *application-level framing* (ALF) and *integrated layer processing* (Clark & Tennenhouse, 1990). In this approach to protocol architecture, the different functions are next to each other, not on top of one another. The Internet multimedia protocol architecture is shown in Figure 1.

As shown in Figure 1, the overall multimedia data and control architecture currently incorporates a set of real-time protocols, which include the real-time transport protocol (RTP) for transporting real-time data and providing QoS feedback, the real-time streaming protocol (RTSP) for controlling delivery of streaming media, the session-announcement protocol (SAP) for advertising multimedia sessions via multicast, and the session-description protocol (SDP) for describing multimedia sessions. In addition, it includes the session-initiation protocol (SIP), which is used to invite the interested parties to join the session. But the functionality and operation of SIP does not depend on any of these protocols. Furthermore, the resource-reservation protocol (RSVP) is designed for reserving network resources. These protocols, together with reliable multicast (Handley, Floyd, Whetten, Kermode, Vicisano, & Luby, 2000), are the underlying support for Internet multimedia applications. While all the protocols above work on top of the IP protocol, the Internet stream protocol, version 2 (ST-II), is an IP-layer protocol that provides end-to-end guaranteed service across the Internet.

The Real Time Transport Protocols: RTP and RTCP

The real-time transport protocol, named as a transport protocol to emphasize that RTP is an end-to-end protocol, is designed to provide end-to-end delivery services for data with real-time characteristics, such as interactive audio and video (Schulzrinne, Casner, Frederick, & Jacobson, 2003). Those services include payload-type identification, sequence numbering, time-stamping, and delivery monitoring. Applications typically run RTP on top of UDP (user datagram protocol) to make use of its multiplexing and checksum services; both protocols contribute parts of the transport protocol functionality. However, RTP may be used with other suitable underlying network or transport protocols. RTP supports data transfer to

Figure 1. Internet protocol architecture for real time applications



nb: For the acronyms, refer to "Key Terms"

multiple destinations using multicast distribution if provided by the underlying network.

RTP consists of two closely-linked parts:

- The real-time transport protocol to carry data that has real-time properties.
- The RTP control protocol (RTCP) to monitor the quality of service and to convey information about the participants in an ongoing session. This functionality may be fully or partially subsumed by a separate session control protocol, which is beyond the scope of this document.

The RTP defines a fixed data-packet header for the set of functions required in common across all the application classes that RTP might support. However, in keeping with the ALF design principle, an extension mechanism is provided to allow the header to be tailored through modifications or additions defined in a profile specification.

RTCP control packets supplement each RTP flow and are periodically transmitted by each participant in an RTP session to all other participants. RTCP performs the following four functions.

1. Provides feedback information to application
2. Identifies RTP source
3. Controls RTCP transmission interval
4. Conveys minimal session-control information

Internet Stream Protocol: ST-II

ST-II has been developed to support efficient delivery of streams of packets to either single or multiple destinations in applications requiring guaranteed data rates and controlled delay characteristics (Schulzrinne, Rao, & Lanphier, 1998). ST-II is an Internet protocol at the same layer as IP (Figure 1). ST-II differs from IP in that every intervening ST-II entity maintains state information for each stream that passes through it. The stream state includes forwarding information, including multicast support for efficiency, and resource information, which allows network or link bandwidth and queues to be assigned to a specific stream. This preallocation of resources allows data packets to be forwarded with low delay, low overhead, and a low probability of loss due to congestion, and thus to support efficient delivery of streams of packets to either single or multiple destinations in applications requiring guaranteed data rates and controlled delay characteristics.

Protocols for Multimedia Session Setup and Control

There are two basic forms of multimedia session-setup mechanisms. These are session advertisement and session invitation. Session advertisements are provided using a session directory, and session invitation (inviting a user to join a session) is provided using a session-invitation protocol such as SIP or packet-based multimedia communication systems standard H.323 (ITU, 1998).

Before a session can be advertised, it must be described using the session-description protocol. SDP describes the content and format of a multimedia session, and the session-announcement protocol is used to distribute it to all potential session recipients.

The Session-Description Protocol

The session-description protocol is used for general real-time multimedia session-description purposes and is purely a format for session description (Handley & Jacobson, 1998). SDP is intended for using different transport protocols as appropriate.

SDP serves two primary purposes. It is a means to communicate the existence of a session, and is a means to convey sufficient information to enable joining and participating in the session. A session description contains the following information.

- Session name and purpose
- The media comprising the session, such as the type of media (video, audio), the transport protocol (RTP/UDP/IP, H.320), and the format of the media (H.261 video, MPEG video)
- Time(s) the session is active.
- Information for receiving those media (addresses, ports, formats, and so on)

The SDP description is announced using the session-announcement protocol.

Session-Announcement Protocol

SAP defines an announcement protocol to be used to assist the advertisement of multicast multimedia conferences and other multicast sessions, and to communicate the relevant session-setup information to prospective participants (Handley, Perkins, & Whelan, 2000). Sessions are described using the session-description protocol, and the session description is the payload of the SAP packet.

SAP supports session announcement and deletion. However, SAP defines no rendezvous mechanism, the SAP announcer is not aware of the presence or absence of any SAP listeners, and no additional reliability is provided over the standard best-effort UDP/IP semantics. A SAP announcer periodically sends an announcement packet to a well-known multicast address and port. A preannounced session can be modified by simply announcing the modified session description. A previously announced session may be deleted.

The announcement contains an authentication header for verifying that changes to a session description or deletion of a session are permitted. It can also be used to authenticate the identity of the session creator.

Session-Initiation Protocol

Not all sessions are advertised, and even those that are advertised may require a mechanism to explicitly invite a user to join a session. The session-initiation protocol is an application-layer control (signaling) protocol that can establish, modify, and terminate multimedia sessions or calls (Rosenberg et al., 2002). SIP can also invite participants to already existing sessions, such as multicast conferences. Media can be added to (and removed from) an existing session. SIP invitations used to create sessions carry session descriptions that allow participants to agree on a set of compatible media types. SIP runs on top of several different transport protocols.

SIP supports five aspects of establishing and terminating multimedia sessions.

- **User location:** Determination of the end system to be used for communication
- **User availability:** Determination of the willingness of the called party to engage in communications
- **User capabilities:** Determination of the media and media parameters to be used
- **Session setup:** "Ringing," establishment of session parameters at both the called and calling party
- **Session management:** Including transferring and terminating sessions, modifying session parameters, and invoking services

Note that SIP is not a vertically integrated communications system. SIP is, rather, a component that can be used with other Internet protocols to build a complete multimedia architecture (e.g., RTP, RTSP).

Controlling Multimedia Servers: RTSP

A standard way to remotely control multimedia streams delivered, for example, via RTP, is the real-time stream-control protocol. Control includes absolute positioning within the media stream, recording, and possibly device control. RSTP is primarily aimed at Web-based media-on-demand

services, but it is also well suited to provide VCR-like controls for audio and video streams, and to provide playback and record functionality of RTP data streams. A client can specify that an RTSP server plays a recorded multimedia session into an existing multicast-based conference, or can specify that the server should join the conference and record it. RTSP acts as a "network remote control" for multimedia servers.

The protocol supports retrieval of media from media servers, the invitation of a media server to a conference, and the addition of media to an existing presentation.

Resource Reservation Protocol

RSVP is an IP resource-reservation setup protocol designed for an integrated-services Internet, and it provides receiver-initiated setup of resource reservations for multicast or unicast data flows that take a significant fraction of the network resources (Braden, Zhang, Berson, Herzog, & Jamin, 1997). Using RSVP, the resources necessary for a multimedia session are reserved, and if no sufficient resource is available, the admission is rejected. The reservations and route setups apply only to packets of a particular session, and RSVP identifies a particular session by the combination of destination address, transport-layer protocol type, and destination port number.

RSVP makes receivers responsible for requesting a specific QoS. A QoS request from a receiver host application is passed to the local RSVP process. The RSVP then carries the request to all the nodes (routers and hosts) along the reverse data path(s) to the data source(s), but only as far as the router where the receiver's data path joins the multicast distribution tree.

Quality of service for a particular data flow is ensured by mechanisms collectively called *traffic control*, and they are a packet classifier, admission control, and "packet scheduler," or some other link-layer-dependent mechanism to determine when particular packets are forwarded.

During reservation setup, an RSVP QoS request undergoes *admission control* and *policy control*. Admission control determines whether the node has sufficient available resources to supply the requested QoS. Policy control determines whether the user has administrative permission to make the reservation.

RSVP does not transport application data, but is rather an Internet control protocol. It uses underlying routing protocols to determine where it should carry reservation requests. As routing paths change, RSVP adapts its reservation to new paths if reservations are in place.

THE FUTURE TRENDS

Technically, there are several types of communication networks that are used to provide multimedia communication services, which include data networks, broadband television networks, integrated-services digital networks, and broadband multiservice networks. The Internet is the most widely deployed data network, and significant advancement has been made to provide communication services beyond its original design. This advancement is centered on QoS provisioning on the Internet so that QoS-sensible applications can be supported across the Internet.

Generally, work on QoS-enabled Internet has led to two distinct approaches: (a) the integrated services architecture (often called *Intserv*) and its accompanying signaling protocol, most importantly, RSVP, and (b) the differentiated services architecture (often called *Diffserv*).

As described in this article, the Internet integrated-services architecture is an extension of the Internet architecture and protocols to provide integrated services, that is, to support real-time as well as the current non-real-time service of IP. This extension is necessary to meet the growing need for real-time service for a variety of new applications, including teleconferencing, remote seminars, telescience, and distributed simulation.

It allows sources and receivers to exchange signaling messages that establish packet classification and forwarding state on each node along the path between them. In the absence of state aggregation, the amount of state on each node scales in proportion to the number of concurrent reservations, which can be potentially large on high-speed links. Integrated services are considered not scalable and best suited in intranet and LAN (local area network) environments.

In the last few years, we have witnessed the development of other models of service differentiation in order to facilitate the deployment of real-time applications, including the differentiated-services architecture, relative priority marking, service marking, label switching, and static per-hop classification. The differentiated-services architecture has received the most attention in the Internet community.

In contrast to integrated services, which use the more stringent and complex quality-of-service approach, the differentiated-services architecture has emerged as an alternative for implementing scalable service differentiation on the Internet. This architecture achieves scalability by aggregating traffic classification states, which are conveyed by means of IP-layer packet marking. Packets are classified and marked to receive a particular per-hop forwarding behavior on nodes along their path. Sophisticated classification, marking, policing, and shaping operations need only be implemented at network boundaries or hosts. Network resources are allocated to traffic streams by service-provisioning policies that govern how traffic is marked and conditioned upon entry to a differentiated-services-capable network, and how that traffic is forwarded within that network. A wide variety of services can be implemented on top of these building blocks (Blake, Black, Carlson, Davies, Wang, & Weiss, 1998). We can foresee that in the future, support for multimedia applications on the Internet, both integrated services and differentiated services, will be developed and deployed, and will complement each other.

CONCLUSION

The Internet has evolved from a provider of simple TCP/IP best-effort service to an emerging integrated-service Internet. This development provides tremendous opportunities for building real-time multimedia applications over the Internet. The protocol stack necessary for classes of quality of service, including real-time applications, is presented as the Internet integrated-service architecture that supports the various service models. The constituent real-time protocols of this architecture are the foundations and the critical support elements for the building of Internet real-time multimedia applications.

REFERENCES

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., & Weiss, W. (1998). *An architecture for differentiated services* [RFC 2475]. Internet Engineering Task Force.

Braden, R., Clark, D., & Shenker, S. (1994). *Integrated services in the Internet architecture: An overview* [RFC 1633]. Internet Engineering Task Force.

Braden, R., Zhang, L., Berson, S., Herzog, S., & Jamin, S. (1997). *Resource ReSerVation protocol (RSVP): Version 1, functional specification* [RFC 2205]. Internet Engineering Task Force.

Clark, D. D., & Tennenhouse, D. L. (1990). Architectural considerations for a new generation of protocols. *Computer Communications Review, 20*(4), 200-208.

Crowcroft, J., Handley, M., & Wakeman, I. (1999). *Internetworking multimedia*. Morgan Kaufmann Publishers.

Department of Commerce. (1998). *The emerging digital economy.*

Handley, M., Floyd, S., Whetten, B., Kermode, R., Vicisano, L., & Luby, M. (2000). *The reliable multicast design space for bulk data transfer* [RFC 2887]. Internet Engineering Task Force.

Handley, M., & Jacobson, V. (1998). *SDP: Session description protocol* [RFC 2327]. Internet Engineering Task Force.

Handley, M., Perkins, C., & Whelan, E. (2000). *Session announcement protocol* [RFC 2974]. Internet Engineering Task Force.

ITU. (1998). *Packet-based multimedia communication systems recommendation H.323.* Geneva, Switzerland: Telecommunication Standardization Sector of ITU.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., et al. (2002). *SIP: Session initiation protocol* [RFC 3261]. Internet Engineering Task Force.

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, J. (2003). *RTP: A transport protocol for real-time applications* [RFC 3550]. Internet Engineering Task Force.

Schulzrinne, H., Rao, A., & Lanphier, R. (1998). *Real time streaming protocol (RTSP)* [RFC 2326]. Internet Engineering Task Force.

Topolcic, C. (Ed.). (1990). *Experimental Internet stream protocol, version 2 (ST-II)* [RFC 1190]. Internet Engineering Task Force.

KEY TERMS

Elastic applications: A kind of network applications that will always wait for data to arrive rather than proceed without it.

HTTP: Short for *hypertext transfer protocol,* the underlying protocol used by the World Wide Web.

IP: Abbreviation of *Internet protocol.* IP specifies the format of packets, also called datagrams, and the addressing scheme. Most networks combine IP with a higher level protocol called transmission-control protocol (TCP), forming TCP/IP networks.

Multicast: Means transmitting a single message to a select group of recipients. A simple example of multicasting is sending an e-mail message to a mailing list.

Protocol: A set of rules that govern the operation of functional units to achieve communication.

Protocol architecture: An organization structure of the communication system, which comprises constituent protocols and the relationships among them.

Real-time applications: One class of applications needs the data in each packet by a certain time, and if the data has not arrived by then, the data is essentially worthless.

RSVP: The Internet standard protocol *resource-reservation protocol* (RSVP).

RTP/RTCP: The Internet standard *real-time transport protocol* (RTP) and *RTP control protocol* (RTCP).

RTSP: The Internet standard *real-time streaming protocol* (RTSP) for controlling delivery of streaming of multimedia data.

SAP: The Internet standard *session-announcement protocol* (SAP) for advertising multimedia sessions via multicast.

SDP: The Internet standard *session-description protocol* (SDP) is used for general real-time multimedia session-description purposes and is purely a format for session description.

Service model: Consists of a set of service commitments. In response to a service request, the network commits to deliver some service.

SIP: The Internet standard *session-initiation protocol* (SIP), a signaling protocol for Internet conferencing, telephony, presence, events notification, and instant messaging.

SMTP: Short for *simple mail-transfer protocol*, a protocol for sending e-mail messages between servers. Most e-mail systems that send mail over the Internet use SMTP to send messages from one server to another. In addition, SMTP is generally used to send messages from a mail client to a mail server.

ST-II: The Internet *stream protocol, version 2* (ST-II), an IP-layer protocol that provides end-to-end guaranteed service across the Internet.

TCP: Abbreviation of *transmission-control protocol*, one of the main protocols in TCP/IP networks. Whereas the IP protocol deals only with packets, TCP enables two hosts to establish a connection and exchange streams of data.

UDP: Abbreviation for *user datagram protocol*, a connectionless protocol that, like TCP, runs on top of IP networks. Unlike TCP/IP, UDP/IP provides very few error-recovery services so is less reliable than TCP, offering instead a direct way to send and receive datagrams over an IP network.

This work was previously published in the Encyclopedia of Information Science and Technology, Vol. 3, edited by Mehdi Khosrow-Pour, pp. 1549-1554, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.2
Location-Based Multimedia Content Delivery System for Monitoring Purposes

Athanasios-Dimitrios Sotiriou
National Technical University of Athens, Greece

Panagiotis Kalliaras
National Technical University of Athens, Greece

INTRODUCTION

Advances in mobile communications enable the development and support of real-time multimedia services and applications. These can be mainly characterized by the personalization of the service content and its dependency to the actual location within the operational environment. Implementation of such services does not only call for increased communication efficiency and processing power, but also requires the deployment of more intelligent decision methodologies.

While legacy systems are based on stationary cameras and operational centers, advanced monitoring systems should be capable of operating in highly mobile, ad-hoc configurations, where overall situation and users roles can rapidly change both in time and space, exploiting the advances in both the wireless network infrastructure and the user terminals' capabilities. However, as the information load is increased, an important aspect is its filtering. Thus, the development of an efficient rapid decision system, which will be flexible enough to control the information flow according to the rapidly changing environmental conditions and criteria, is required. Furthermore, such a system should interface and utilize the underlying network infrastructures for providing the desired quality of service (QoS) in an efficient manner.

In this framework, this article presents a *location-based multimedia content delivery system* (LMCDS) for monitoring purposes, which incorporates media processing with a decision support system and positioning techniques for providing the appropriate content to the most suitable users, in respect to user profile and location, for monitoring purposes. This system is based on agent technology (Hagen & Magendanz, 1998) and aims to promote the social welfare, by increasing

the overall situation awareness and efficiency in emergency cases and in areas of high importance. Such a system can be exploited in many operational (public or commercial) environments and offers increased security at a low cost.

SERVICES

The LMCDS provides a platform for rapid and easy set up of a monitoring system in any environment, without any network configurations or time-consuming structural planning. The cameras can be installed in an ad hoc way, and video can be transmitted to and from heterogeneous devices using an intelligent decision support system (IDSS) according to the user's profile data, location information, and network capabilities.

Users can dynamically install ad-hoc cameras to areas where the fixed camera network does not provide adequate information. The real-time transmission of still images or video in an emergency situation or accident to the available operational centers can instantly provide the necessary elements for the immediate evaluation of the situation and the deployment of the appropriate emergency forces. This allows the structure of the monitoring system to dynamically change according to on-the-spot needs.

The IDSS is responsible for overviewing the system's activity and providing multimedia content to the appropriate users. Its functionality lies in the following actions:

- Identifying the appropriate user or group of users that need access to the multimedia content (either through user profile criteria or topological criteria).
- Providing the appropriate multimedia content in relevance to the situation and the location.
- Adapting the content to the user's needs due to the heterogeneity of the users devices—that is, low bit rate video to users with portable devices.

The LMCDS can evaluate users' needs and crisis events in respect to the topological taxonomy of all users and provide multimedia content along with geographical data. The location information is obtained through GPS or from GPRS through the use of corresponding techniques (Markoulidakis, Desiniotis, & Kypris, 2004). It also provides intelligent methodologies for processing the video and image content according to network congestion status and terminal devices. It can handle the necessary monitoring management mechanisms, which enable the selection of the non-congested network elements for transferring the appropriate services (i.e., video streaming, images, etc.) to the concerned users. It also delivers the service content in the most appropriate format, thus allowing the cooperation of users equipped with different types of terminal devices.

Moreover, the LMCDS provides notification services between users of the system for instant communication in case of emergency through text messaging or live video feed.

All of the above services outline the requirements for an advanced monitoring system. The LMCDS functionality meets these requirements, since it performs the following features:

- Location-based management of the multimedia content in order to serve the appropriate users.
- Differentiated multimedia content that can be transmitted to a wide range of devices and over different networks.
- Lightweight codecs and decoders that can be supported by devices of different processing and network capabilities.
- IP-based services in order to be transparent to the underlying network technology and utilize already available hardware and operating systems platforms.
- Intelligent delivery of the multimedia content through the LMCDS in order to avoid increased traffic payload as well as information overload.

- Diverse localization capabilities through both GPS and GPRS, and generation of appropriate topological data (i.e., maps) in order to aid users.

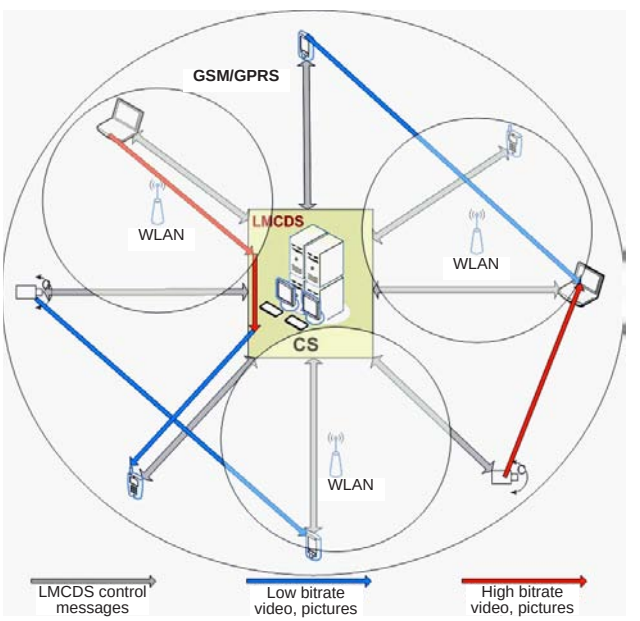
However, the system architecture enables the incorporation of additional location techniques (such as WLAN positioning mechanisms) through the appropriate, but simple, development of the necessary interfaces with external location mechanisms.

In order to describe the above services in a more practical way, a short list of available options and capabilities of the system is given below. The target group of the LMCDS consists of small to medium businesses that seek a low-cost solution for monitoring systems or bigger corporations that need an ad hoc extension to their current system for emergency cases and in order to enhance their system's mobility. Even though the users of the system consist mainly of security staff and trained personnel that are in charge of security, the system's ease of use, low user complexity, and device diversity allow access even to common untrained users.

The system offers a range of capabilities, most of which are summarized in Figure 1, such as:

- User registration and authentication.
- User profile (i.e., device, network interface).
- Location awareness:
  - User is located through positioning techniques.
  - User is presented with appropriate topographical information and metadata.
  - User is aware of all other users' locations.
  - User can be informed and directed from a Center of Security (CS) to specified locations.
- Multimedia content:
  - Video, images, and text are transmitted to user in real time or off-line based on situation or topological criteria.

Figure 1. System functionality and services



  - User can provide feedback from his device through camera (laptop, PDA, smart phone) or via text input.
  - Content is distributed among users from the CS as needed.
- Ad hoc installation of cameras that transmit video to CS and can take advantage of wireless technology (no fixed network needed).
- Autonomous nature of users due to agent technology used.

LMCDS ARCHITECTURE

The LMCDS is designed to distribute system functionality and to allow diverse components to work independently while a mass amount of information is exchanged. This design ensures that new users and services can be added in an ad hoc manner, ensuring future enhancements and allowing it to support existing monitoring systems.

Multi-agent systems (MASs) (Nwana & Ndumu, 1999) provide an ideal mechanism for implementing such a heterogeneous and sophisticated distributed system in contrast to traditional software technologies' limitations in communication and autonomy.

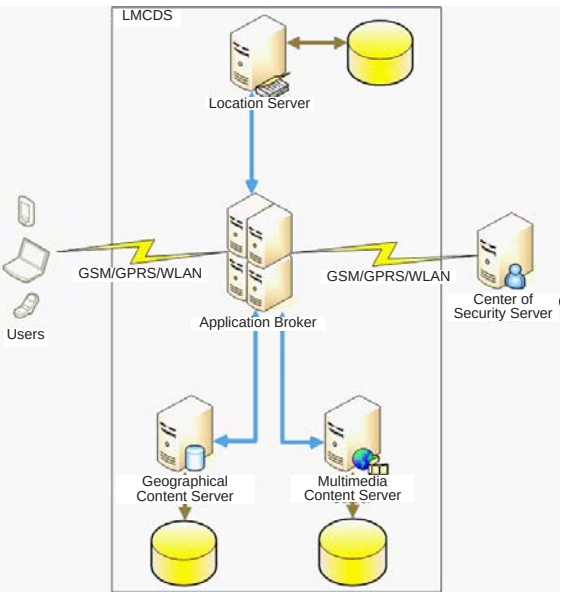
The system is developed based on MAS and allows diversified agents to communicate with each other in an autonomous manner, resulting in an open, decentralized platform. Tasks are being delegated among different components based on specific rules, which relate to the role undertaken by each agent, and information is being composed to messages and exchanged using FIPA ACL (FIPA, 2002a). An important aspect for the communication model is the definition of the content language (FIPA, 2002b). Since LMCDS targets to a variety of devices, including lightweight terminals, the LEAP language (Berger, Rusitschka, Toropov, Watzke, & Schlichte, 2002) of the JADE technology has been exploited.

In addition to the security mechanisms supported by the underlying network components, the JADE platform offers a security model (Poggi, Rimassa, & Tomaiuolo, 2001) that enables the delegation of tasks to respective agent components by defining certificates, and ensures the authentication and encryption of TCP connections through the secure socket layer (SSL) protocol (http://www.openssl.org/).

The general architecture of the LMCDS is shown in Figure 2. The platform is composed of different agents offering services to the system which are linked by an application broker agent, acting as the coordinator of the system. These agents are the location server, the center of security server, the application broker, the geographical content server, and the multimedia content server. The latter two are discussed in later sections in more detail, while a brief description of the functionality of the others is given as follows.

The location server agent is responsible for the tracking of all users and the forwarding of location-based information to other components. The information is gathered dynamically and

Figure 2. Architecture overview



kept up-to-date according to specific intervals. The intelligence lies in the finding of the closest users to the demanded area, not only in terms of geographical coordinates, but also in terms of the topology of the environment. More information on location determination is given in a following section.

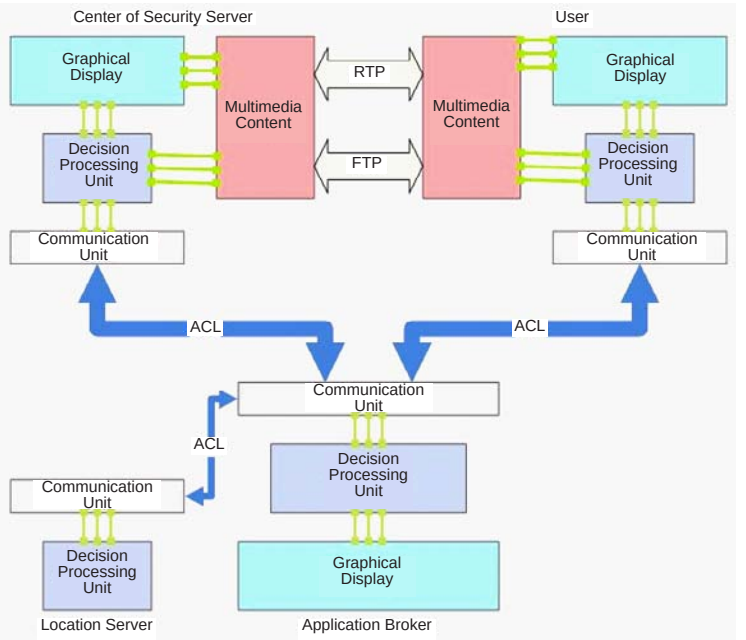
The center of security server agent monitors all users and directs information and multimedia content to the appropriate users. It is responsible for notifying users in emergency situations, and also performs monitoring functions for the system and its underlying network.

The User agent components manage information, including the transmission and reception of image or video, the display of location information, critical data, or other kinds of displays. They are in charge of several user tasks, such as updating users' preferences and location, decoding the response messages from the application broker, and performing service level agreement (SLA) (ADAMANT, 2003) monitoring functionalities. The user agent can reside in a range of devices, since it is Java based.

Finally, the application broker agent acts as a mediator that provides the interface between all of the components. It is responsible for prioritizing user requests according to users' profiles, performing authentication functions, and acting as a facilitator for SLA negotiations. It also coordinates the communication process between users in order for the multimedia content to be delivered in the appropriate format according to the user's processing and network capabilities, and also the network's payload.

A closer look into the components of each agent, along with the interaction between them, is shown in Figure 3. Each agent is composed of three main components: the graphical display, which is responsible for user input and information display; the communication unit, in charge of agent communication through ACL messages; and the decision processing unit, which processes all received information. In addition, the user and the CS include a multimedia content component

Figure 3. Agent components



for the capture, playback, and transmission of multimedia data through RTP or FTP channels.

MULTIMEDIA PROCESSING AND ENCODING

Video/Image Formats

One of the novelties is the ability of the system to perform real-time format conversions of the image or video data and transmit to several heterogeneous recipients, ranging from large PC servers to small personal devices like PDAs or mobile phones.

The LMCDS enables the adoption and support of different video formats, according to the partial requirements of the user terminals and the available networks status. The most commonly used is the M-JPEG (http://www.jpeg .org/index.html) format. It was preferred over other common video formats suitable for wireless networks like MPEG (http://www. m4if.org/) and H.263 (http://www.itu.int/rec/recommendation.asp), which provide higher compression ratio, because using them can require intense processing power both for the encoder and the decoder. Also, frames in MPEG or H.263 streams are inter-related, so a single packet loss during transmission may degrade video quality noticeably. On the contrary, M-JPEG is independent of such cascaded-like phenomena, and it is preferable for photo-video application temporal compression requirements for smoothness of motion.

It is a lossy codec, but the compression level can be set at any desired quality, so the image degradation can be minimal. Also, at small data rates (5-20Kbps) and small frame rates, M-JPEG produces better results than MPEG or H.263. This is important, as the photos or video can be used as clues in legal procedures afterwards, where image quality is more crucial than smooth motion. Another offering feature is the easy extraction of JPEG (http://www.jpeg .org/index.html) images from video frames.

The video resolution can be set in any industry-standard (i.e., subQCIF) or any other resolution of width and height dividable of 8, so the track is suitable for the device it is intended for. Video is streamed directly from the camera-equipped terminals in a peer-to-peer manner. Transmission rates for the video depend on the resolution and the frame rate used. Some sample rates are given in Table 1.

Apart from M-JPEG, another set of video formats have been adopted, such as H263 and MPEG-4. The development of these formats enables the testing and evaluation of the LMCDS, based on the network congestion and the current efficiency of the supported video formats and the crisis situations in progress. This means that for a specific application scenario, the encoding with M-JPEG format can lead to better quality on the user terminal side, while the MPEG 4 format can be effective in cases that the network infrastructure is highly loaded, so the variance in bit rate can keep the quality in high values.

Image compression is JPEG with resolution of any width and height dividable of 8. For the real-time transmission of video stream, the real-time transfer protocol (RTP, http://www.ietf.org/rfc/rfc3550.txt) is used, while for stored images and video tracks, the file transfer protocol (FTP) is used.

Table 1. Output video formats for the application

Resolution	Frame Rate	Suitable Network	Trans. Rate (kbps)	Target Device
160 x 120	1	GPRS WLAN	2 – 3	Smartphone PDA, PC
160 x 120	5	GPRS WLAN	10 – 15	Smartphone, PDA,PC
232 x 176	2	GPRS WLAN	10 – 15	PDA,PC
320 x 240	5	WLAN	30 – 40	PC
320 x 240	15	WLAN	90 – 120	PC
640 x 480	5	WLAN	45 – 55	PC

Video Processing

It is important to point out that the output video formats can be produced and transmitted simultaneously with the use of the algorithm shown in Figure 4.

Note that the image/video generator can be called several times for the same captured video stream as long as it is fed with video frames from the frame grabber. So, a single user can generate multiple live video streams with variations, not only in frame rate and size, but also in JPEG compression quality, color depth, and even superimpose layers with handmade drawings or text. The algorithm was implemented in Java with the use of the JMF API (http://java.sun.com/products/java-media/jmf/).

For a captured stream at a frame rate of n frames per second, the frame grabber component extracts from the raw video byte stream n/A samples per second, where A is a constant representing the processing power of the capturing device. Depending on how many different qualities of video streams need to be generated, m Image/Video Generator processes are activated, and each process i handles K_i fps. The following relationship needs to be applied:

$$K_i = B_i * n / A * m \text{ , where}$$

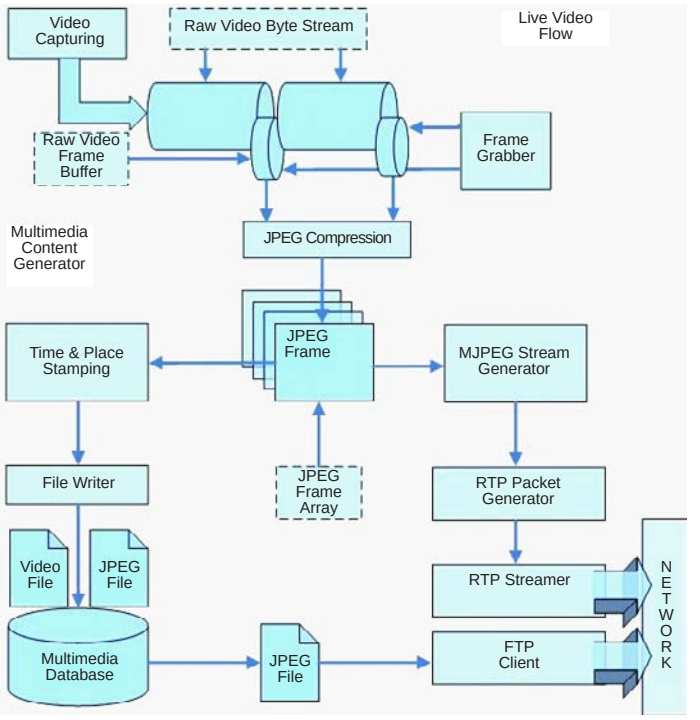
$$B_i = (K_i * A * m) / n \text{ and}$$

$$\sum B_i = m$$

There is also a latency of $m*C_i$ seconds per video stream, where C_i depends on the transcoding time of the image to each final format.

So, this tendency of the LMCDS to keep the frame rate low is inevitable due to this sampling process. However, using low frame rates is quite common in surveillance systems. It also allows long hours of recording, where video size is optimal when minimum both for storage and transmission, and is less demanding in process-

Figure 4. LMCDS video processing overview



ing power for use with video players running on small devices.

GEOGRAPHICAL CONTENT DELIVERY

Positioning Methods

The LMCDS uses the following techniques for locating the users' positions inside the served environment.

- *GPS (Global Positioning System)* is a global satellite system based on a group of non-geostatic satellites in middle altitude orbit (12,000 miles). The GPS-enabled devices have the ability to locate their position with a high degree of accuracy by receiving signals from at least six satellites.
- A GSM/GPRS subscriber can be located upon request, depending on the received power from adjacent cells. More information about these mechanisms can be found in Markoulidakis et al. (2004).
- The *Ekahau* position engine (http://www.ekahau.com/pdf/EPE_2.1_datasheet.PDF) is designed to locate 802.11b (at present) wireless LAN users positions in the indoor environment. In the context of radio resource management, the software could apply the access point's radio coverage map in the indoor environment. It uses a calibration method. Initially it measures a set of sample points' radio strength. Based on these sample points, the engine can estimate a client WLAN station's approximate location, given an arbitrary combination of signal strengths.

Geographical Database

The geographical database of the LMCDS is storing information about the positions of the users that are registered in the system. The information is obtained regularly by scheduled queries. When the users perform service request messages to the system, their position coordinates are automatically retrieved (through any of the above methods), so their location in the geographical DB is also updated. Special importance has been given to the ability of the system to serve queries about the relative position of its users and estimation of distances. So, the scheduled queries can be of the following types:

- Find my 3 nearest users and send them a video.
- Estimate the time that I need to get to Building A.
- Find the users closest to point B and send pictures to those that operate in GPRS network and high-quality video to those in UMTS or WLAN.
- When user C enters a specified area, send him a message.

The user is displayed visual information in the form of maps to its device. The map consists of raster data—that is, the plan of the area and also several layers of metadata, showing points of interest, paths, as well as the position of relative users.

FUTURE TRENDS

Future steps involve the exploitation of video streaming measurements for providing guaranteed QoS of the video content to the end user, as well as the better utilization of the available radio resources. Furthermore, the incorporation of new trends in video streaming in conjunction with a markup language for multimedia content, such as MPEG-7 or MPEG-21, can offer a higher level of personalized location-based services to the end user and are in consideration for future development.

CONCLUSION

This article presented a location-based multimedia content system enabling real-time transfer of multimedia content to end users for location-based services. Based on the general architecture of multi-agent systems, it focused on fundamental features that enable the personalization of the service content and the intelligent selection of the appropriate users for delivering the selected content.

REFERENCES

Berger, M., Rusitschka, S., Toropov, D., Watzke, M., & Schlichte, M. (2002). Porting distributed agent-middleware to small mobile devices. *Proceedings of the Workshop on Ubiquitous Agents on Embedded, Wearable, and Mobile Devices,* Bologna, Italy.

FIPA (Foundation for Intelligent Physical Agents). (2002a). *FIPA ACL message structure specification.* SC00061G.

FIPA. (Foundation for Intelligent Physical Agents). (2002b). *FIPA SL content language specification.* SC00008I.

Hagen, L., & Magendanz, T. (1998). Impact of mobile agent technology on mobile communication system evolution. *IEEE Personal Communications, 5*(4).

IST ADAMANT Project. (2003). *SLA management specification.* IST-2001-39117, Deliverable D6.

Markoulidakis, J. G., Desiniotis, C., & Kypris, K. (2004). Statistical approach for improving the accuracy of the CGI++ mobile location technique. *Proceedings of the Mobile Location Workshop , Mobile Venue '04.*

Nwana, H., & Ndumu, D. (1998). A perspective on software agents research. *The Knowledge Engineering Review, 14*(2).

Poggi, A., Rimassa, G., & Tomaiuolo, M. (2001). Multi-user and security support for multi-agent systems. *Proceedings of WOA 2001 Workshop,* Modena, Italy.

KEY TERMS

Agent: A program that performs some information gathering or processing task in the background. Typically, an agent is given a very small and well-defined task.

Application Broker: A central component that helps build asynchronous, loosely coupled applications in which independent components work together to accomplish a task. Its main purpose is to forward service requests to the appropriate components.

IDSS: Intelligent decision support system.

LMCDS: Location-based multimedia content delivery system.

MAS: Multi-agent system.

Media Processing: Digital manipulation of a multimedia stream in order to change its core characteristics, such as quality, size, format, and so forth.

Positioning Method: One of several methods and techniques for locating the exact or relative geographical position of an entity, such as a person or a device.

This work was previously published in the Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 381-386, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.3
Provisioning of Multimedia Applications Across Heterogeneous All–IP Networks

Michail Tsagkaropoulos
University of Patras, Greece

Ilias Politis
University of Patras, Greece

Tasos Dagiuklas
Technical Institute of Messolonghi, Greece

Stavros Kotsopoulos
University of Patras, Greece

INTRODUCTION

With the opening of the telecommunication market and the emergence of low-cost and heterogeneous wireless access technologies, it is envisaged that next-generation network and service providers will not only vary in the deployed access technology but also in their business models and structures. Such providers will differ from large providers such as the current telecom providers offering multiple services and covering large geographical areas, down to small providers offering certain services such as conferencing or messaging only or covering small geographical areas such as a coffee shop or a shopping mall. Further, while in the current networking environment, a home provider of a user is usually represented by a large telecom provider; in such a heterogeneous environment, any trustworthy entity such as an application provider, a banking entity, or a credit card provider that is capable of authenticating the user and marinating his usage profile can act as a home provider. Towards this vision this article discusses the issues that concern the establishment of multimedia applications across heterogeneous networks.

NEXT-GENERATION NETWORKS AND THE ALL-IP CONVERGENCE

Convergence of heterogeneous wireless technologies over a broadband IP core network will allow mobile subscribers to access a new variety of services, over a variety of access networks and by using a variety of devices. This integration will be realized on the network access with devices able to hand off across heterogeneous wireless access technologies, service delivery, and availability (Dagiuklas & Velentzas, 2003). There is no industry consensus on what next-generation networks will look like but, as far as the next-generation networks are concerned, ideas and concepts include:

- Transition to an "All-IP" network infrastructure
- Support of heterogeneous access technologies (e.g., UTRAN, WLANs, WiMAX, xDSL, etc.)
- VoIP substitution of the pure voice circuit switching
- Seamless handovers across both homogeneous and heterogeneous wireless technologies
- Mobility, nomadicity, and QoS support on or above the IP layer
- Provisioning of triple-play services creating a service bundle of unifying video, voice, and Internet
- Home networks opening new doors to the telecommunication sector and network providers
- Unified control architecture to manage application and services
- Convergence among network and services.

HETEROGENEOUS MULTIMEDIA NETWORKS AND SERVICES

Vision

In order to allow seamless communication and roaming in a heterogeneous wireless environment, one needs to provide an efficient way for coupling multimedia service provisioning, access with fast-handover schemes, and establishing trust relations among service providers and users. This necessitates the provisioning of a framework for establishing security and trust relations among network operators, service providers, and mobile users, allowing thereby smooth roaming among different administrative domains/networks and seamless provisioning of multimedia services. Such a vision infrastructure is illustrated in Figure 1.

This infrastructure will support the dynamic establishment of trust relations between independent providers (e.g., foreign and home providers) in a distributed manner over hybrid IPv4 and IPv6 networks (Salkintzis, 2004). Moreover, it will provide the required enhancements for providing secure interconnection among different heterogeneous networks, establishing user-provider trust relations, and the necessary means for authenticating users in foreign domains and exchanging their profiles in a secure manner. This would thereby enable users to roam to foreign networks and use the provided services in these networks without affecting their privacy. Finally, to support the smooth and fast handover, efficient and secure context exchange mechanisms will be provided, allowing users to roam among different providers without having to explicitly re-authenticate themselves and establish new trust relations.

It is envisioned that the NGN architecture will be based on packet-based technologies. The most important part of NGN is the division of network functionality into many distributed functions, which fall into the following categories (Dagiuklas et al., 2005):

Figure 1. Inter-domain framework

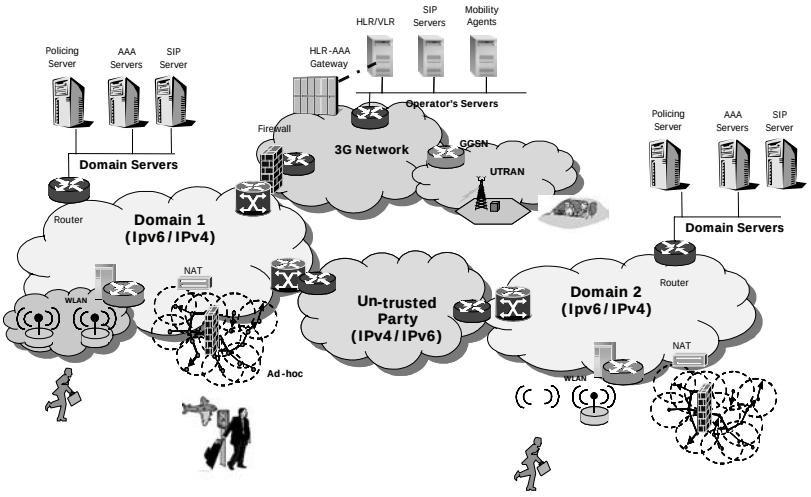
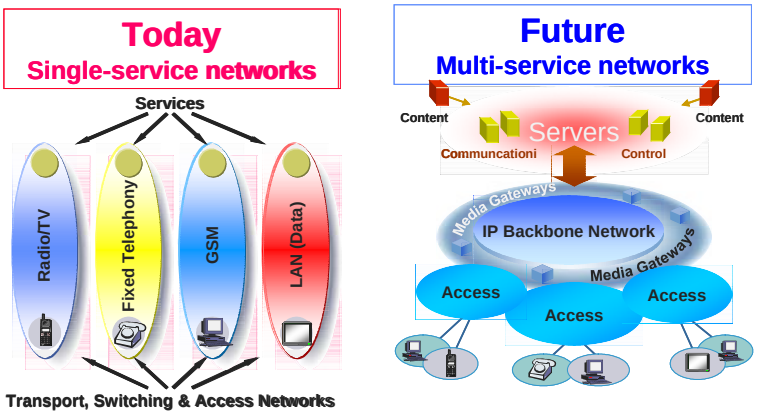


Figure 2. NGN architecture migration (Nokia, 2001)



1. Control, management, and signaling, which provides the intelligence needed for user control of the connection. This intelligence is distributed.
2. Access, routing, switching, and transport, which provides the functions needed for transporting information between end users and other network elements.
3. Convergence with existing legacy networks (PSTN, SS7, mobile networks).

Figure 2 illustrates the transitions from the current scene towards NGN.

To achieve the goal of enabling flexible roaming between different network and services providers in a secure and chargeable manner between providers of different technologies and business models, this article will be addressing the following requirements (Velentzas & Dagiuklas, 2005):

- Secure and scalable inter-domain AAA infrastructure for mobile user (secure roaming)
- User access security
- Application level security

- Concepts for security context transfer while the user hands over among heterogeneous wireless technologies.

INTER-PROVIDER RELATIONSHIP

The establishment of heterogeneous networks aims to import substantial services for the customer through the specification and realizing of infrastructure that enable dynamic trust relation among different providers in a secure and scalable manner. Particular emphasis should be given to roaming business models while the mobile users move across different administrative domains. The possibly very large number of service and network providers as well as home providers makes the establishment of static trust relations between all of those providers tedious and non-economical. Therefore, mechanisms and solutions must be defined that will allow users to roam from one provider (either network operator or service provider or a mixture of both) to the other without having to explicitly subscribe to the services of each provider or assuming some pre-established trust relation between the different providers.

Dynamic Trust Relations

Most current solutions for enabling roaming of users among different network providers assume the existence of pre-established trust relations between the involved partners. Two providers that would like to enable user roaming must be registered at this broker. Currently, the number of network providers is relatively small, with those providers having a large scale in terms of the number of supported users and covered geographical area. Further, the definition of a home provider might change as well, allowing for example a banking unit to act as the home provider of a user and authenticate his identity. Registering all possible providers at a single trust broker would surely not scale. This would then lead to having different brokers each maintaining trust relations to only a small number of providers. To allow for roaming between any two entities, the foreign network needs to be able to discover the home provider of the user and which trust brokers could be used for establishing a trust relation with the home provider. The experience gained for private key distribution architectures shows that for such schemes to work, they need to be distributed. The work to be done here would include specifying a distributed trust infrastructure and investigating its applicability and performance with the increasing number of providers in NGN.

Identity Management

Identity management is best defined as "those IT and business processes, organizations, and technologies that are applied to ensure the integrity and privacy of identity and how it translates to access." This results in its effective use as a crucial element of IT security infrastructure. Recently, identity management has been elevated within many IT organizations to be a formal program consideration by several business drivers.

The Liberty framework, developed by the Liberty Alliance project (www.projectliberty.org), comprises an identity management architecture focused on the establishment of secure, consistent, and manageable business relationships and interactions over the Internet. This architecture is realized along circles of trusts embracing affiliated identity providers and service providers. The Liberty framework further defines adequate schemas and metadata for identity management, along with the required bindings to Web-based middleware (SOAP) services.

Current roaming solutions require close interaction between the home and foreign providers in order to gain access to user identity profiles. 3GPP solutions necessitate for example that any signaling messages sent by the user be routed through the home network. This not only increases the set-up delay for establishment of communication

sessions, but also complicates the offering of local services in foreign networks. To overcome this problem the home and foreign networks need to be able to exchange user profiles that allow the foreign network to make the decision of whether or not to grant a user access to certain services locally. However, while doing this the privacy of the user's data and profile need to be guaranteed, and the foreign provider should only be given information that is needed for authorizing the user and providing him with the necessary access rights.

Policy-Driven, Configurable, and Programmable AAA Infrastructures

The exact behavior of AAA infrastructure might vary considerably depending on its location (intra- or inter-domain), functionality (broker, proxy, home, or foreign), supported features (QoS, mobility, multimedia), provider policy, load balancing architectures, or security requirements for example. To ease the replacement of RADIUS protocol-based servers to diameter-based AAA servers, server and network providers need to have powerful and yet simple interfaces for programming and customizing those servers (Nakhjiri & Nakhjiri, 2005). While Diameter-based AAA technologies are increasingly being considered as the basis for AAA in NGN, there will always be providers relying on proprietary or older solutions. To still be able to communicate with such providers and exchange AAA information with them, some translation mechanisms need to be used to cover the gap.

Reliable and Secure Intra- and Inter-Provider AAA Infrastructure

While attacking a certain server of access router by generating a lot of useless traffic might render part of the network or a certain service useless, attacking the AAA infrastructure would render the complete network useless. Mounting an attack on a trust broker would make any roaming between the networks impossible. Currently, the AAA infrastructure in PSTN networks is not very vulnerable simply because the end systems that are allowed to connect to the network are "dumb." However, in the open environment of NGN mounting attacks on an AAA server will become easier. Any system can start a large number of authentication requests and occupy thereby not only a substantial share of the network bandwidth but also of the processing resources available to the AAA servers.

Another aspect to be considered here is the reliability of the AAA infrastructure in the face of various failure situations such as software or hardware failures of the AAA servers themselves or the links connecting those servers to the networks. The following points need to be dealt with in future work:

- Identification of attack and exploitation possibilities on AAA infrastructures
- Dynamic detection of attacks on AAA infrastructure and devising of mechanisms for defending and protecting against those attacks and reducing their effects
- Specification and realization of fail-over mechanisms for AAA servers

Secure Multimedia Service Access

In the literature of node cooperation enforcement, the proposed solutions can be subdivided into two main categories: trade-based schemes and reputation-based schemes. In trade-based schemes, a node that provides some service to a peer node (e.g., packet forwarding) is rewarded by either another immediate service in exchange or some monetary token that he can later use to buy services from another node. In reputation-based schemes each node keeps a reputation metric for other nodes it deals with and provides services only to nodes that exhibit good reputation.

In all reputation-based mechanisms for cooperation enforcement, each node in the network performs two distinct functions: rating the behavior of neighboring nodes and using these ratings to adjust its own behavior towards them. Rating the conformance of neighboring nodes to a given network protocol is an operation that depends on the specific protocol and network architecture. For instance, in single-channel MANETs, rating the packet forwarding service provided by a node's neighbors is simply performed through monitoring of the common channel. However, in clustered MANETs, which use different channels in each cluster and bridge nodes to relay packets between clusters (such as Bluetooth scatter nets), a node cannot receive the transmissions of all of his neighbors. Hence, a different technique for rating the forwarding services provided by them is needed. Similarly, rating the conformance to a neighborhood discovery protocol or a Medium Access protocol is fundamentally different than rating packet forwarding.

On the other hand, a cooperation reinforcing reputation mechanism can be easily adapted to use such behavior ratings independently of the rated service. A crucial task for this mechanism is to distinguish between perceived and actual non-cooperative behavior. For example an MT might receive a bad cooperation rating because of link failure or mobility. Misbehaving MTs might also choose to misbehave in a probabilistic way in order to evade detection. If erroneously perceived misbehavior is permitted with a certain probability, then the detection of intentional misbehavior is reduced to an estimation problem.

Secure Service Discovery

The service discovery may be performed using a hierarchical multi-tier approach based on several tiers. A service manager (SM) discovers and manages the services in its corresponding tier and interacts with its upper-tier SM. All the available public and private services provided by foreign nodes, directly attached to the network, or provided by any other kind of infrastructure network will be discovered, subject to authentication and authorization.

In the service discovery architecture, the protocols used for communication are some common service discovery protocols, such as UPnP, Bluetooth SDP, and JXTA. At the highest level, the objective is to create an environment where message-level transactions and business processes can be conducted securely in an end-to-end fashion. There is therefore a need to ensure that messages are secured during transit, with or without the presence of intermediate nodes. There may also be a need to ensure the security of the data in storage. The requirements for providing end-to-end security for the service discovery are summarized as follows:

- **Secure service registration and deregistration:** Service registration and deregistration corresponds to service information management by the SMs. Only authorized service providers should be allowed to register and deregister a service from the repository. Meanwhile, it is important to maintain the integrity and confidentiality of the registered services in the service registration and deregistration process. Other types of attacks should be mitigated such as message replay and spoofing. Efficient message authentications through signatures or keyed hash-functions are suitable countermeasures for these attacks.
- **Secure authorization:** Authorized service discovery is needed to control the discoverable services by each entity. On the one hand, users may be authorized to perform a service discovery, but based on their set of credentials, they may only have a controlled visibility of available services. On the other hand, the discovered service must be genuine and trustworthy. Finally, the anonymity of the entity performing the service discovery

and confidentiality of the process should also be ensured.

- **Secure Service Delivery/Provisioning:** After a service has been found, it should be securely delivered by the mutual authentication between the recipient of the service and the service provider. Delivery confidentiality and service integrity must also be met so that the genuine and authentic service is only delivered to the intended recipient(s). As for secure service discovery, anonymity with reference to the location and identity can also be required.
- **Dependability:** Dependability can be defined as the property of a system, which always honors any legitimate requests by authorized entities. It is violated when an attacker succeeds in denying service access to legitimate users (e.g., by exhausting all the available resources through the DoS attack).
- **Service Access Control:** Future architectures should contain an AAA component that assigns each service a security profile describing the required security and trust level of the user in order to access the service (Sisalem & Kuthan, 2004)). Profiles determine the access rights for users and contain authorized users, credentials needed, certificates, subscription to groups, and so forth. Profile management can be done by the node, offering the particular service (de-centralized approach) and SM as the service advertiser when evaluating the decision to respond to queries (centralized approach). Using both centralized and de-centralized management methods, it will protect the various access networks (3G, WLAN/WMAN, PAN/PN) from unauthorized use and will keep the management procedure fast and reliable.

Context-Aware Security

One of the basic requirements for B3G is to support seamless mobility, such that the user does not perceive any delay or interruption of service in heterogeneous networks. To provide seamless mobility, handover between networks of either heterogeneous technologies or different administrative domains must be smooth and secure. Because each network may deploy its own security mechanisms that are incompatible with others, seamless handover imposes certain restrictions for maintaining the same security level and minimum delay.

Context transfer aims to minimize the impact of certain transport/routing/security-related services on the handover performance (Loughney, Nakhjiri, Perkins, & Koodli, 2004). When a mobile node (MN) moves to a new subnet, it needs to maintain services that have already been established at the previous subnet. Such services are known as 'context transfer candidate services', and examples of these services include QoS policy, AAA profile, IPsec state, header compression, session maintenance, and so forth. Re-establishing these services at the new RAN will require a considerable amount of time for the protocol exchanges, and as a result time-sensitive real-time traffic will suffer during this time. Alternatively, context transfer candidate services state information can be transferred, for example, from the previous RAN to the new RAN so that the services can be quickly re-established. A context transfer protocol will result in a quick re-establishment of context transfer candidate services at the new domain. It would also contribute to the seamless operation of multimedia application streams and could reduce susceptibility to errors. Furthermore, re-initiation to and from the mobile node will be avoided, hence wireless bandwidth efficiency will be conserved (Georgiades, Dagiuklas, & Tafazolli, 2006).

Context transfer requirements should consider the following:

1. **Context management and user control:** In many activities in the research community, context information is considered to be important to support usage, operation, and management of heterogeneous wireless network (3G, WLANs, emerging 4G) and service provided by these networks.
2. **Secure context transfer of security information in handovers between heterogeneous access technologies or network types:** When a handover occurs, timing constraints may forbid performing a full new access procedure, including authentication and key agreement. Instead, the security context may be transferred between points of attachment in the network which trust each other. The precise nature of the transferred security context must be specified, and the security for the discovery of points of attachment and of the transferred context need to be further studied and researched, especially during vertical handover.
3. **Secure context adaptation of security information in handovers between heterogeneous access networks:** It may not be sufficient to merely transfer the security context in a handover, but the security context may need to be adapted according to the new environment. For instance, the IP address in an IPsec Security Association may change, or different cryptographic mechanisms or schemes to protect communication traffic may be used.

MULTIMEDIA PROVISIONING

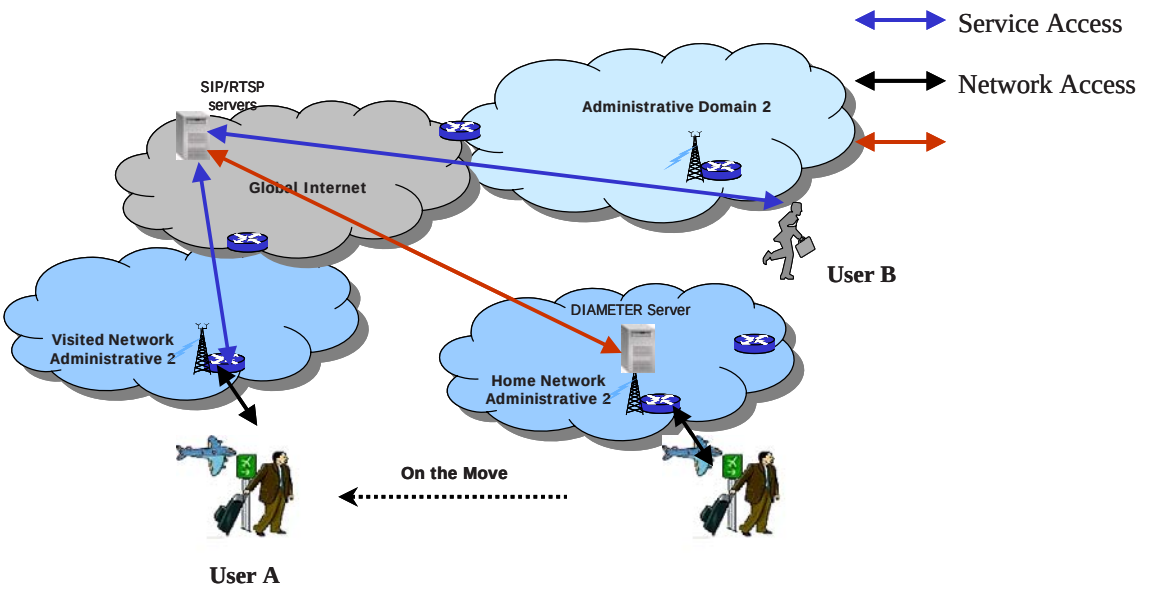
The implementation of demanding services, such as real-time applications and streaming multimedia, in mobile environments using wireless connections has attracted considerable attention over the last few years. Efficient mobility management is considered to be one of the major factors towards seamless provision of multimedia applications across heterogeneous networks (Dagiuklas et al., 2005).

In order to cope with the requirements imposed by the NGN architecture, the IP multimedia subsystem (IMS) has been specified by 3GPP as a comprehensive service framework for both basic calling and enhanced multimedia services (3GPP, 2004). The IMS is the key enabler in the mobile world for providing rich multimedia services to the end users. The IMS enables complex IP-based multimedia sessions to be created with guaranteed QoS for each media component, ensuring that multimedia sessions will be able to reserve the resources they need and are authorized to use in order to perform satisfactorily. The IMS does not standardize any applications, only the service capabilities required to build various services. As a result, real-time and non-real-time multimedia services can easily be integrated over a common IP-based transport. The functions supported by the IMS include quality of service control, interworking and roaming, service control, and multiple network access. Session initiation protocol (SIP) was chosen as the session control protocol for IMS, which is based on HTTP and uses all service frameworks developed for HTTP. In addition diameter was chosen to be the AAA protocol in the IMS.

Taking into account a mixed environment with both fixed and mobile users and the currently standardized protocols SIP and diameter, several extensions for a complete and integrated security framework with a well-defined identification mechanism and a fully defined intercommunication method with AAA architecture are needed (Camarillo & Garcia-Martin, 2004). Multimedia provisioning across heterogeneous networks should consider the following aspects:

- **Application level security:** SIP security mechanisms can be embodied within SIP body message including encryption (enables the possibility to ensure privacy) and AAA.

Figure 3. AAA and SIP interworking scenario



- **SDP security and media security extensions:** A number of SDP extensions have been motivated by SIP-based applications, and these need to be accommodated in SDP, which now supports the use of the SRTP/SRTCP protocol. The definition of the SRTP protocol is still in draft phase and under discussion. Features associated with key management attributes need to be included (not just for SIP) and so may need to be general mechanisms to signal security capabilities.
- **SIP identity integrity:** The identity information asserted by the sender of a request is the 'From' header containing an URI (like 'sip: ipolitis@ee.upatras.gr') and an optional display name (like "Ilias") that identifies the originator of the request.
- **SIP privacy:** The privacy problem is further complicated by proxy servers (also referred to in this document as "intermediaries" or "the network") that add headers of their own, such as the record-route and via headers.
- **End-to-End QoS:** Current approaches focus on QoS control in the access part, but the QoS control among service providers is still an open issue (Farkas et al., 2006).

CONCLUSION

In conclusion this article relates the functionalities and AAA infrastructure in order to support the dynamic establishment of trust relations between independent providers in a secure and distributed manner. The support of NGN networks involves research work on vast areas ranging from mobility, quality of service (QoS), roaming models, security, and integration with current networks. Towards the establishment of heterogeneous networks and services, solutions have been presented for supporting the provision of multimedia services over heterogeneous networks, benefiting from the availability of standardized solutions (IMS, Diameter, SIP, etc.) for supporting the operators' needs and solving issues of heterogeneity. Finally, implementing the envisaged conditions, the users' freedom to roam between different networks and use any locally available services in secure and satisfactory manner is increased.

REFERENCES

Camarillo, G., & Garcia-Martin, M. A. (2004). *The 3G IP multimedia subsystem (IMS).* New York: John Wiley & Sons.

Dagiuklas, T., & Velentzas, S. (2003). *3G and WLAN interworking scenarios: Qualitative analysis and business models.* IFIP HET-NET03, Bradford, UK.

Dagiuklas, T., Gatzounas, D., Theofilatos, D., Sisalem, D., Rupp, S., Velentzas, R., et al. (2002). Seamless multimedia services over all-IP network infrastructures: The EVOLUTE approach. *Proceedings of the IST Summit 2002* (pp. 75-78).

Dagiuklas, T., Politis, C., Grilli, S., Bigini, G., Rebani, Y., Sisalem, D., et al. (2005). Seamless multimedia sessions and real-time measurements across hybrid 3G and WLAN networks. *International Journal of Wireless and Mobile Computing,* (4th Quarter).

Farkas, K., Wellnitz, O., Dick, M., Gu X., Busse, M., Effelsberg, W., et al. (2006). Real-time service provisioning for mobile and wireless networks. *Elsevier Computer Communications, 29*(5), 540-550.

Georgiades, M., Dagiuklas, T., & Tafazolli, R. (2006). Middlebox context transfer for multimedia session support in all-IP networks. *Proceedings of the ACM Conference IWCMC,* Vancouver, Canada.

Kingston, K., Morita, N., & Towle, T. (2005). NGN architecture: Generic principles, functional architecture and implementation. *IEEE Communications Magazine,* 49-56.

Loughney, J., Nakhjiri, M., Perkins, C., & Koodli, R. (2004). *Context transfer protocol.* Internet Draft, *draft-ietf-seamoby-ctp-08.txt.*

Nakhjiri, M., & Nakhjiri, M. (2005). *AAA and network security for mobile access* (pp. 1-23). New York: John Wiley & Sons.

Salkintzis, A. (2004). Interworking techniques and architectures for WLAN/3G integration towards 4G mobile data networks. *IEEE Wireless Communications,* (June), 50-61.

Sisalem, D., & Kuthan, J. (2004). Inter-domain authentication and authorization mechanisms for roaming SIP users. *Proceedings of the 3rd International Workshop on Wireless Information Systems,* Porto, Portugal.

3GPP. (2004). *IP multimedia subsystem version 6.* 3G TS 22.228.

Velentzas, S., & Dagiuklas, T. (2005). *Tutorial: 4G/wireless LAN interworking.* IFIP HET-NET 2005, Ilkley, UK.

KEY TERMS

Authentication Authorization Accounting (AAA): Provides the framework for the construction of a network architecture that protects the network operator and its customers from attacks and inappropriate resource management and loss of revenue.

Diameter: An AAA protocol for applications such as network access or IP mobility. It is a base protocol that can be extended in order to provide AAA services to new access technologies; it is intended to work in both local and roaming AAA situations**.**

IP Multimedia Subsystem (IMS): Provides a framework for the deployment of both basic calling and enhanced multimedia services over IP core.

Liberty Alliance: Broad-based industry standards consortium developing suites of specifications defining federated identity management and Web services communication protocols that are suitable for both intra-enterprise and inter-enterprise deployments.

Quality of Service (QoS): The probability of the telecommunication network meeting a given traffic contract, or in many cases used informally to refer to the probability of a packet succeeding in passing between two points in the network.

MANET: Mobile ad-hoc network.

Next Generation Networking (NGN): A broad term for a certain kind of emerging computer network architectures and technologies which generally describes networks that natively encompass data and voice (PSTN) communications, as well as (optionally) additional media such as video.

Remote Authentication Dial-In User Service (RADIUS): An AAA protocol for applications such as network access or IP mobility.

Session Initiation Protocol (SIP): A protocol developed by the IETF MMUSIC Working Group and proposed standard for initiating, modifying, and terminating an interactive user session that involves multimedia elements such as video, voice, instant messaging, online games, and virtual reality. It is one of the leading signaling protocols for voice over IP.

Service Manager (SM): A server that discovers and manages services in its corresponding tier and interfaces with its upper-tier SM.

Triple Play: A term for the provisioning of the three services—high-speed Internet, television (video-on-demand or regular broadcasts), and telephone service—over a single broadband wired or wireless connection.

This work was previously published in the Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 796-803, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.4
Adaptive Narrative Virtual Environments

Karl Steiner
University of North Texas, USA

ABSTRACT

Narratives are an important method of human-to-human communication. Combining the power of narrative with the flexibility of virtual environments (VEs) can create new and innovative opportunities in education, in entertainment, and in visualization. In this chapter, we explore the topic of narrative VEs. We describe the characteristics and benefits of narrative VEs, review related work in VEs and in computer-generated narrative, and outline components of an architecture for managing narrative VEs. We present the current status of our work developing such an architecture and conclude by discussing what the future of narrative VEs may hold.

INTRODUCTION

Storytelling is a significant method of human-to-human communication. We tell stories to share ideas, to convey emotions, and to educate. As new communication technologies have become available we have employed them in our storytelling, allowing us to reach wider audiences or to tell stories in new ways. Narratives were among the initial and most popular content types as books, radio, movies, and television were introduced to the public. VEs may provide the next technological advancement for presenting narratives. Consider the following examples.

Illustration 1

An inexperienced army specialist enters an immersive simulator in order to practice skills such as identifying and overcoming threats in an urban environment. The specialist encounters a hostile unit on Main Street, but chooses to retreat to a safe location. No longer sure of her ability to handle the hostile unit, the officer continues to explore the environment, but avoids Main Street. Aware that the specialist is no longer able to interact with the hostile unit on Main Street, the environment attempts to surprise the officer with a new hostile unit near her current location. This time she is better prepared. Rather than retreat, she engages and subdues the hostile unit.

Illustration 2

A child, nervous about an upcoming surgery, is given an opportunity to explore a virtual hospital. The child follows a patient about to undergo similar surgery, sees all the preparations that take place, and becomes familiar with the various people who are involved in the procedure. While he does not see the actual surgery take place, the environment makes sure the child notices that there are other children waiting for surgery and that they are all experiencing many of the same emotions that he is. He is reassured after his exploration — not just with his knowledge of the procedure, but that his fears and concerns are normal.

Illustration 3

A young adult begins interacting with a new mystery game set in a virtual environment. In the initial interactions, the player continually chooses to explore the environment and look for clues, rather than chase and subdue suspects. Noting the players interests, the game adjusts the plotline to emphasize puzzles rather than conflicts.

These examples illustrate three powerful, yet underutilized, potentials of VEs. The first attribute is that events occur not randomly, but according to a plan based on scenario or narrative goals. The second attribute is that the system may recognize and respond to user goals and interests in addition to the given scenario goals. And the third attribute is that the system chooses when, where, and how to present events in order to best meet the user and scenario goals.

In this chapter, we will explore the topic of narrative VEs. We begin by describing the characteristics and benefits of narrative VE. We review related work that has influenced the current state of both VE and computer-generated narrative and outline components of an architecture for managing narrative VE. We describe the current status of our work developing such an architecture and conclude by discussing what the future of narrative VE may hold.

BACKGROUND

Narrative VE

Narrative VEs are an emerging multimedia technology, integrating storytelling techniques into VEs. VEs and virtual reality have been defined in different ways. For this chapter, we will define a VE as a computer-generated 3D space with which a user can interact. The concept of narrative also has different meanings in different contexts. In our work, we consider a narrative to be a sequence of events with a conflict and resolution structure (a story). So, for us, a narrative VE is an immersive, computer-generated 3D world in which sequences of events are presented in order to tell a story.

Not all VEs are narrative, nor are VEs the only way of presenting computer-based narratives. Most VEs do not have an explicitly narrative structure. While most VEs include events of some sort, they do not provide the plot-driven selection and ordering of events that separates stories from simulation. For example, a driving simulator might model the roadways of a city, complete with cars, pedestrians, and traffic signals. Events in this simulation-oriented VE might include a traffic light changing from green to red, a pedestrian successfully crossing an intersection, or a collision between two cars.

In a simulation, the rules that govern when and how events occur are based primarily on attributes of the world and the object within it. In our driving VE, the stoplight may cycle from green to yellow to red every 30 seconds. Pedestrians cross roads successfully when there is no traffic. A collision may occur whenever the paths of two objects intersect at the same point in space and time. While simulation rules are adequate to describe object-object interactions and can even be used to model very complex environments (such as flying a 747 or driving through city traffic), they do not typically address broader goals or plans of the users or authors of the environment. The rules and random probabilities of a simulator do

not take into account how an event will impact the environment or the user.

In a narrative VE, however, the occurrences of at least some key events are based on storytelling goals. For example, in a narrative VE designed to illustrate the dangers of jaywalking, the stoplight may turn green as soon as the user's car approaches the intersection. The pedestrian may wait to start crossing the intersection until the user's car enters the intersection. While these events may be similar to those in a simulator (traffic signal changes color, pedestrian crosses the street), the timing, placement, and representation of these events is done in such a manner as to communicate a particular story. This is what differentiates a simulation from a narrative.

Benefits of Narrative VE

Combining the power of narrative with the power of VEs creates a number of potential benefits. New ways of experiencing narratives as well as new types of VE applications may become possible. Existing VE applications may become more effective to run or more efficient to construct.

Narrative VE can extend our traditional ideas of narrative. Unlike traditional narrative media, narrative VE can be dynamic, nonlinear, and interactive. The events in a narrative VE would be ultimately malleable, adapting themselves according to a user's needs and desires, the author's goals, or context and user interaction. A narrative VE might present a story in different ways based on a user profile that indicates whether the user prefers adventure or romance. A user could choose to view the same event from the viewpoint of different characters or reverse the flow of time and change decisions made by a character. A narrative VE might adapt events in order to enhance the entertainment, the education, or the communication provided by the environment. And a narrative VE might adapt events in order to maintain logical continuity and to preserve the suspension of disbelief (Murray, 1998).

These narrative capabilities could also open up new application areas for VE. Training scenarios that engage the participant, provide incrementally greater challenges, and encourage new forms of collaboration could become possible (Steiner, 2002). Interactive, immersive dramas could be created where the user becomes a participant in the story (Laurel, 1991). Presentations of temporal or spatial data could be self-organizing, arranging and rearranging the data based on different perspectives or emphases (Gershon, 2002).

While it is possible to add narrative elements to an otherwise simulation-oriented VE, the process can be difficult for designers, and the results less than satisfying for users. The most common application area for simulation-oriented VEs with narrative elements is computer games. A review of the current state of the art in 3D narrative games is instructive in considering the challenges VE designers face.

Many computer games now include narrative elements, though the quality and integration of the narratives vary widely. One of the most common narrative elements is a full screen video (FSV) cut-scene that advances the storyline. These typically take place after a player has completed a level or accomplished some other goal. Other devices include conversations with non-player characters (NPCs) that are communicated as actual speech, as text, or as a FSV. These conversations are typically scripted in advance, with the only interactivity coming from allowing the user to choose from prepared responses.

Even with the best-produced narrative games, the narratives are generally tightly scripted and linear, providing little or no opportunity for influencing the plot of the story. Since the environment remains unaware of narrative goals, the designer must take measures to force users into viewing narrative events. This usually means either eliminating interactivity and forcing users to watch FSVs; or keeping interactivity, but eliminating choices so that the user must ultimately proceed down a particular path or choose a particular

interaction. These solutions are unsatisfying for many users and time consuming for designers to construct.

A fundamental issue behind the limited interactivity is the tension between the user's desire to freely explore and interact with the environment and the designer's goals of having users experience particular educational, entertaining, or emotional events. Advances in visualization technology have exacerbated the issue. VEs are now capable of modeling more objects, behaviors, and relationships than a user can readily view or comprehend. Techniques for automatically selecting which events or information to present and how to present it are not available, so VE designers must explicitly specify the time, location, and conditions under which events may occur. As a result, most VEs support mainly linear narratives or scenarios (if they support them at all).

Related Work

While there has been substantial research on the application of certain intelligent techniques to VEs, there has been little attention to adaptive interfaces for narrative VE. Still, in realizing an architecture for adaptive narrative VE, many lessons can be drawn from work in related areas. These include adaptive interfaces, intelligent agents in VE, automated cinematography in VE, narrative multimedia, and narrative presentation in VE.

Adaptive interfaces are those that change based on the task, work context, or the needs of the user. Often, the successful application of an adaptive interface requires simultaneous development of a user modeling component. A user modeling system allows the computer to maintain knowledge about the user and to make inferences regarding their activities or goals.

Adaptive interfaces allow computer applications to be more flexible, to deal with task and navigation complexity, and to deal with divers user characteristics and goals. Adaptive interfaces can be applied to almost any human/computer interaction. For example, the immense popularity of the WWW has lead to a corresponding amount of research on adaptive hypermedia. Similarly, adaptive techniques based on user models can also be employed in VEs.

Automated cinematography is a form of adaptivity particularly for VEs. Camera controllers or automated cinematographers are designed to free users and designers from explicitly specifying views. Simple camera controllers may follow a character at a fixed distance and angle, repositioning in order to avoid occluding objects or to compensate for user actions such as running. More complex controllers seek to model cinematography convention (He, 1996), to dynamically generate optimal camera angles to support explicitly stated user viewing or task goals (Bares, 1999), or to link camera control to a model of the ongoing narrative (Amerson, 2001).

Over the years, researchers have created various systems for dynamically generating textual narratives. One of the earliest systems, Talespin, created fable-like stories (Meehan, 1980). These narratives exhibited a conflict-resolution structure driven by characters seeking to achieve their goals; but the text was frequently repetitive, and the stories were not particularly engaging. Universe (Lebowitz, 1984) created a dynamic textual narrative based on a simulation of characters in a soap opera. While such a setting provides rich dramatic content, the narrative was again dictated by characters pursuing goals, rather than being dictated by a plot. In more recent work (Calloway, 2001) researchers have created the Author Architecture and the Storybook application, a narrative generation system capable of creating real-time fairy tale text almost on a par with human generated text.

A natural extension of the character driven textual narratives are character-driven VEs. Several VE systems have been built which allow users to interact with lifelike, believable, or emotional agents (Hayes-Roth, 1996). As with the

character-driven textual systems, the results of interacting with these agents may be interesting or believable; but they may not exhibit an overall narrative coherence.

There has also been work on VE environments that seek to provide such coherence. In the OZ Architecture for Dramatic Guidance (Weyhrauch, 1997) a director gives instructions to actors in a VE, seeking to maximize the dramatic potential of a situation. Mimesis (Young, 2001) includes a director that works to detect and prevent potential plot conflicts due to user activities. The Dogmatics project (Galyean, 1995) attempts to direct user attention by manipulating the environment.

AN ARCHITECTURE FOR NARRATIVE VE INTELLIGENT INTERFACE

Building on this prior work, we are developing an architecture for adaptive narrative in VEs. This architecture combines components common to adaptive systems and VE systems and adds data and control structures capable of representing and reasoning about narratives. The architecture includes the following components: Adaptation Manager, Narrative Manager, User Modeler, and a 3D Engine. While other VE systems devoted to narrative presentation have addressed elements of VE architecture, we believe that our approach to adapting events represents a unique contribution to the field. Our narrative VE architecture is illustrated in Figure 1.

3D Engine and World State

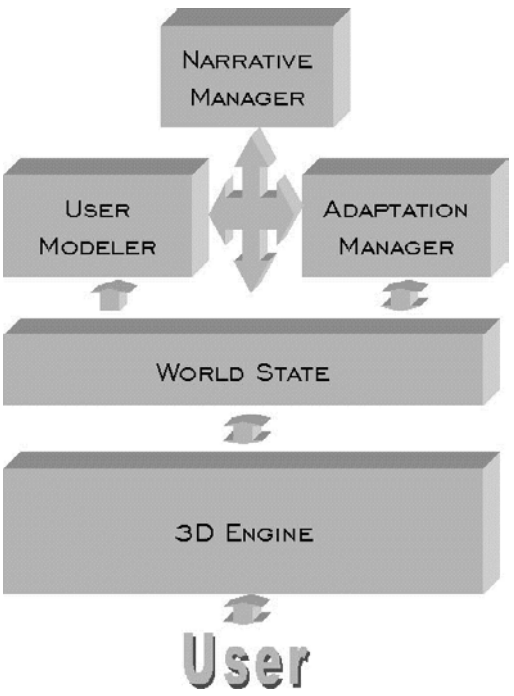
Graphic VEs of the type we have been describing (as opposed to textual VEs such as MUDs and MOOs) rely on a 3D Engine to render the real-time images that make up the visual presentation of the environment. While there are many relevant research topics related to presentation of 3D images, our research focuses primarily upon technologies related to interaction, adaptivity, and narrative. Accordingly, rather than develop our own 3D engine, we have taken advantage of the many powerful VE environments that are now available to researchers and commercial developers. Like a growing number of academic researchers, we have found that the ease of development and flexibility of the current suite of game engines meets our development needs (Lewis, 2002). In particular, we are using the A5 3D game engine from Conitec.

The world state contains information about the object within the world, such as its location and its current status. This information may be maintained within the 3D engine or may be managed by an external data store.

User Modeler

User modeling describes the ability of a system to make observations about a user and to use

Figure 1. Adaptive narrative VE architecture



these observations to adapt the experience to better meet the user's wants or needs. For a narrative VE system, the user model could receive information such as a user profile, a history of user activities within the VE, and information about the activities of others with the VE. Using this information, the User Modeler could provide information regarding the user beliefs about the current story or which aspects of the story appear interesting to the user.

In order to develop reliable inferences, the User Modeler would require a variety of input. This input could include information regarding user information and attributes (e.g., age, occupation, computer experience, hobbies and interests, etc.). The user model would also receive a record of activities performed by the user within the environment. This activity log would be more than just a list of specific interactions (e.g., movement to coordinates X, Y, Z, interaction with object Q, etc.), but would also provide contextual information used to establish the significance of interactions (e.g., moved farther away the threatening dog, interacted with the door to unlock the secret room, etc.). Using this information, the system could generate various "beliefs" about the user, such as the user's goals (the user seeks recreation and intellectual stimulation), the users interests within the narrative VE (the user is seeking treasure and wealth), what the user believes about the narrative or the VE (the user believes there is a treasure in the hidden room or the user believes that the dog is hostile and will attack), and whether the user falls into a recognized "type" (the user is an "explorer" rather than a "fighter").

Narrative Manager

In order to present a story, a VE requires some control mechanism for scheduling and presenting the events that make up the narrative. While simulation-oriented VE have rules for triggering or scheduling events, they do not take into account the narrative purpose of the events, instead focusing on some other guiding structure such as game physics. In the simplest narrative VEs (such as many recent computer games), narrative events are "canned" FSV pieces that occur when certain conditions are met, such as completion of a level, or the solution of a puzzle. In a more robust narrative VE, there would be more variety in the type of events, in the scheduling of events, and in the presentation of events. For example, in addition to or instead of FSV, events might be more contextually inserted into the environment itself. Instead of a pre-recorded FSV showing one character talking to another at a particular location, a real-time animation of the characters can be dynamically created showing their conversation at any location. Instead of narrative events occurring only between levels, narrative events might occur anytime or anywhere within the environment, with the timing, placement, and form of the event being dictated by the storytelling goals of the system (see Adaptation Manager). The basic outline of the narrative (or the plot of the story) could be predefined by a human author or could be generated dynamically by the computer.

In our current implementation, the outline of the story is expressed through narrative events (see Figure 2). Other interactions may take place in the system, but the narrative events are those that have significance to the story, have corresponding triggers, and have their presentation controlled by the Adaptation Manager.

Adaptation Manager

Given an event and contextual information, the system should be capable of modifying the event to maintain plot and logical continuity and to help the user and author achieve their narrative goals. For the author, these goals may include communicating an idea, teaching a lesson, or conveying an emotion. For the user, goals might include entertainment, education, relaxation, or excitement. As input, the Adaptation Manager would use information from the Narrative Manager (narrative events), the User Modeler (user goals or beliefs), and the world state (object loca-

Figure 2. Event data structures

NARRATIVE EVENT		
SLOT	**DESCRIPTION**	
Name	Name of Event	
Location	Location where event should occur	
Time	When event should occur	
Participants	Actors who are involved in event	
Actions	Set of actions that make up the event	
Presentation	The preferred form of communicating the actions, e.g. direct observation, indirect observation, hearing, intermediated, etc.	
Side Effect	Facts or flags that should be set as a result of this event	
NARRATIVE EVENT TRIGGER		
CRITERIA	**DESCRIPTION**	**EXAMPLES**
Time	Events may need to occur at a specific or relative time or in a specific sequence.	`Session-ends` should occur 20 minutes after `session begins;` `Car-breaks-on-ice` should occur immediately after `car-accelerates-past-user`
Location	Events may need to occur when actors (human or computer) arrive at or leave certain locations	`Stoplight-turns-red` should occur when user arrives at corner of main and center
Other	Events may need to occur when other conditions (indicated by flags, often set as a result of a side-effect associated with an event)	`Car-stops` should occur when `car-out-of-gas` is true
ACTION		
SLOT	**DESCRIPTION**	
Actor	Actor or object who is to perform the action	
Action	Specific primitive action that the actor performs (e.g. animation, movement, function, etc.)	
Object (optional)	Actors or object to or on whom the actor performs the action	
Other Parameters (optional)	Any other parameters necessary to specify the Action	

tions). Given a particular event suggested by the Narrative Manager, the Adaptation Manager will select time, place, and representation based on event constraints, narrative goals, and the state of the world.

CURRENT STATUS

Development is proceeding in phases. We have completed a pilot adaptive narrative VE that includes a limited set of events, a scripted narrative, and an active Presentation Manager. We are in the process of conducting user studies comparing user comprehension and satisfaction both with and without presentation management. This feedback will guide us in extending the supported set of events and the types of adaptation supported by the Adaptation Manager. We have also begun work on the User Modeler and plan to turn our attention next to dynamic narrative generation. The following sample interaction provides an

example of some of the current types of adaptation supported by our system.

Sample Interaction

In order to test the Adaptation Manager, our event representation scheme, and our overall framework, we have created a sample narrative. The narrative was constructed to include enough events and complexity to exercise a variety of adaptation techniques. The narrative includes multiple characters, a plot consisting of multiple events (including several events that may occur simultaneously), and the events take place at multiple locations in a 3D VE.

Event: A hungry rabbit eats the last of his carrots and goes looking for something else to eat.
Actions: Animation of rabbit eating carrots. Sounds of rabbit eating carrots.
Adaptations: Amplify sound of rabbit eating carrots (to draw attention to rabbit and convey idea that someone is eating something). (Figure 3)
Event: The rabbit notices a snowman's carrot nose and steals it.
Actions: Animation and sounds of rabbit stealing carrot nose.
Adaptations: Delay actions until user is present at the characters' location (so that user may view and hear this key event). (Figure 4)

Event: The snowman follows the rabbit to the lake, but the rabbit has already eaten the carrot. The only item he can find is a dead fish, so he uses that as a temporary replacement.
Actions: Sounds of snowman and rabbit discussing options. Animation of snowman picking up fish and using it as nose.
Adaptations: If the user was not present to see the event, the rabbit talks to himself and recounts the event within the hearing distance of the user. (Figure 5)

FUTURE TRENDS

If the public's past and current appetite for narrative and interactive technology are any guides, then continued development and use of new narrative technologies such as adaptive VEs is likely to be rapid. This development could usher in a generation of more sophisticated and personalized computer training, communication, and entertain-

Figure 3.

Figure 4.

ment applications, providing more opportunities for people to interact with multimedia technologies. While some may have concerns that such systems could isolate users by providing them with replacements for human interaction, we feel that some of the greatest opportunities for such technologies would be in creating novel and powerful new ways for people to interact with each other. Collaborative narrative VEs could bring users together to explore stories and situations, even if the users were in physically remote locations. In addition to experiencing narratives in new ways, advances in input and output technologies may also allow us to experience narrative in new places. Using augmented or mixed reality technology, a VE could be overlaid on top of the real world, allowing virtual characters and objects to appear side-by-side with real objects. Such technologies would allow us to experience stories anywhere and to incorporate any location into part of an ongoing personal narrative.

CONCLUSION

The application of AI techniques in narrative generation and adaptation to multimedia technologies such as VEs can create new types of experiences for users and authors. We have explored an architecture that supports not only the presentation of a single narrative, but an adaptive capability that would allow narratives to be customized for different users, different goals, and different situations. The development of this pilot application has demonstrated the feasibility of this architecture, and we plan to continue to extend and enhance the capabilities of our system.

Figure 5.

REFERENCES

Amerson, D., & Kime, S. (2001). Real-time cinematic camera control for interactive narratives. In *The Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment.* AAAI.

Bares, W. H., & Lester, J. C. (1999). Intelligent multi-shot visualization interfaces for dynamic 3d worlds. *Proceedings of the 1999 International Conference on Intelligent User Interfaces* (pp. 119-126). New York: ACM Press.

Callaway, C. B., & Lester, J. C. (2001). Narrative prose generation. *Proceedings of the 17th International Joint Conference on Artificial Intelligence* (pp. 241-248). Morgan Kaufmann.

Galyean, T. (1995). *Narrative guidance of interactivity.* Doctoral dissertation, Massachusetts Institute of Technology.

Gershon, N., & Page, W. (2002). What storytelling can do for information visualization. *Communications of the ACM*, *44*(8), 31-37.

Hayes-Roth, B. & van Gent, R. (1996). Story-making with improvisational puppets and actors. *Stanford Knowledge Systems Laboratory Report KSL-96-05.*

He, L., Cohen, M. F., & Salesin, D. H. (1996). The virtual cinematographer. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques.* New York: ACM Press.

Laurel, B. (1991). *Computers as Theatre*. NY: Addison-Wesley.

Lebowitz, M. (1984). Creating characters in a story-telling universe. *Poetics, 13,* 171-194.

Lewis, M., & Jacobson, J. (2002). Introduction to the special issue on game engines in scientific research. *Communications of the ACM, 45*(1), 27-31.

Meehan, J. (1980). *The Metanovel: Writing Stories by Computer.* New York: Garland Publishing.

Murray, J. (1998). *Hamlet on the Holodeck*. Cambridge, MA: MIT Press.

Steiner, K. E., & Moher, T. G. (2002). Encouraging task-related dialog in 2d and 3d shared narrative workspaces. *Proceedings of the 4th International Conference on Collaborative Virtual Environments.* New York: ACM Press.

Weyhrauch, P. (1997). *Guiding interactive drama.* Doctoral dissertation, Carnegie Mellon University. Technical Report CMU-CS-97-109.

Young, M. R. (2001). An overview of the mimesis architecture: Integrating intelligent narrative control into an existing gaming environment. *The Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment.*

This work was previously published in Computer Graphics and Multimedia: Applications, Problems and Solutions, edited by J. DiMarco, pp. 72-85, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.5
Semantically Driven Multimedia Querying and Presentation

Isabel F. Cruz
University of Illinois, Chicago, USA

Olga Sayenko
University of Illinois, Chicago, USA

ABSTRACT

Semantics can play an important role in multimedia content retrieval and presentation. Although a complete semantic description of a multimedia object may be difficult to generate, we show that even a limited description can be explored so as to provide significant added functionality in the retrieval and presentation of multimedia. In this chapter we describe the Delaunay^{View} that supports distributed and heterogeneous multimedia sources and proposes a flexible semantically driven approach to the selection and display of multimedia content.

INTRODUCTION

The goal of a semantically driven multimedia retrieval and presentation system is to explore the semantics of the data so as to provide the user with a rich selection criteria and an expressive set of relationships among the data, which will enable the meaningful extraction and display of the multimedia objects. The major obstacle in developing such a system is the lack of an accurate and simple way of extracting the semantic content that is encapsulated in multimedia objects and in their inter-relationships. However, metadata that reflect multimedia semantics may be associated with multimedia content. While metadata may not be equivalent to an ideal semantic description, we explore and demonstrate its possibilities in our proposed framework. Delaunay^{View} is envisioned as a system that allows users to retrieve multimedia content and interactively specify its presentation using a semantically driven approach.

Delaunay^{View} incorporates several ideas from the earlier systems Delaunay (Cruz & Leveille, 2000) and Delaunay^{MM} (Cruz & James, 1999). In the Delaunay^{View} framework, multimedia content is stored in autonomous and heterogeneous sources

annotated with metadata descriptions in resource description framework (RDF) format (Klyne & Carroll, 2004). One such source could be a database storing scientific aerial photographs and descriptions of where and when the photographs were taken. The framework provides tools for specifying connections between multimedia items that allow users to create an integrated virtual multimedia source that can be queried using RQL (Karvounarakis et al., 2002) and keyword searches. For example, one could specify how a location attribute from the aerial photo database maps to another location attribute of an infrared satellite image database so that a user can retrieve images of the same location from both databases.

In Delaunay^{View}, customizable multimedia presentation is enabled by a set of graphical interfaces that allow users to bind the retrieved content to presentation templates (such as *slide sorters* or *bipartite graphs*), to specify content layout on the screen, and to describe how the dynamic visual interaction among multimedia objects can reflect the semantic relationships among them. For example, a user can specify that aerial photos will be displayed in a slide sorter on the left of the workspace, satellite images in another slide sorter on the bottom of the workspace, and that when a user selects a satellite image, the aerial photos will be reordered so that the photos related to the selected image appear first in the sorter.

In this paper we describe our approach to multimedia querying and presentation and focus on how multimedia semantics can be used in these activities. In "Background" we discuss work in multimedia presentation, retrieval, and description; we also introduce concepts relating to metadata modeling and storage. In "A Pragmatic Approach to Multimedia Presentation", we present a case study that illustrates the use of our system and describe the system architecture. In "Future Work" we describe future research directions and summarize our findings in "Conclusions."

BACKGROUND

A multimedia presentation system relies on a number of technologies for describing, retrieving and presenting multimedia content. XML (Bray et al., 2000) is a widely accepted standard for interoperable information exchange. MPEG-7 (Martinez, 2003; Chang et al., 2001) makes use of XML to create rich and flexible descriptions of multimedia content. Delaunay^{View} relies on multimedia content descriptions for the retrieval and presentation of content, but it uses RDF (Klyne & Carroll, 2004) rather than XML. We chose RDF over XML because of its richer modeling capabilities, whereas in other components of the Delaunay^{View} system we have used XML (Cruz & Huang, 2004).

XML specifies a way to create structured documents that can be easily exchanged over the Web. An XML document contains *elements* that encapsulate data. *Attributes* may be used to describe certain properties of the elements. Elements participate in hierarchical relationships that determine the document structure. XML Schema (Fallside, 2001) provides tools for defining elements, attributes, and document structure. One can define typed elements that act as building blocks for a particular schema. XML Schema also supports inheritance, namespaces, and uniqueness.

MPEG-7 (Martínez, 2003) defines a set of tools for creating rich descriptions of multimedia content. These tools include *Descriptors*, *Description Schemes (DS)* (Salembier & Smith, 2001) and the *Description Definition Language (DDL)* (Hunter, 2001). MPEG-7 descriptions can be expressed in XML or in binary format. Descriptors represent low-level features such as texture and color that can be extracted automatically. Description Schemes are composed of multiple Descriptors and Description Schemes to create more complex descriptions of the content. For example, the MediaLocator DS describes the location of a multimedia item. The MediaLocator

is composed of the MediaURL descriptor and an optional MediaTime DS: the former contains the URL that points to the multimedia item, while the latter is meaningful in the case where the MediaLocator describes an audio or a video segment. Figure 1 shows an example of a MediaLocator DS and its descriptors RelTime and Duration that, respectively, describe the start time of a segment relative to the beginning of the entire piece and the segment duration.

The Resource Description Framework (RDF) offers an alternative approach to describing multimedia content. An RDF description consists of statements about *resources* and their *properties*. An RDF resource is any entity identifiable by a URI. An RDF statement is a triple consisting of subject, predicate, and object. The subject is the resource about which the statement is being made. The predicate is the property being described. The object is the value of this property. RDF Schema (RDFS) (Brickley & Guha, 2001) provides mechanisms for defining resource classes and their properties. If an RDF document conforms to an RDF schema (expressed in RDFS), resources in the document belong to classes defined in the schema. A class definition includes the class name and a list of class properties. A property definition includes domain — the subject of the corresponding RDF triple — and range — the object. The RDF Query Language (RQL) (Karvounarakis et al., 2002) is a query language for RDF and RDFS documents. It supports a select-from-where structure, basic queries and iterators that combine the basic queries into nested and aggregate queries, and generalized path expressions.

MPEG-7 Description Schemes and Delaunay^{View} differ in their approach to multimedia semantics. In MPEG-7, semantics are represented as a distinct description scheme that is narrowly aimed at narrative media. The Semantic DS includes description schemes for places, objects, events, and "agents" — people, groups, or other active entities — that operate within a narrative world associated with a multimedia item. In Delaunay^{View} any description of multimedia content can be considered a semantic description for the purposes of multimedia retrieval. Delaunay^{View} recognizes that, depending on the application, almost any description may be semantically valuable. For example, an aerial photo of the arctic ice sheet depicts some areas of intact ice sheet and others of open water. Traditional image processing techniques can be applied to the photo to extract light and dark regions that represent ice and open water respectively. In the climate research domain, the size, shape, and locations of these regions constitute the semantic description of the image. In MPEG-7 however, this information will be described with the StillRegion DS, which does not carry semantic significance.

Beyond their diverse perspectives on the nature of the semantics that they incorporate, MPEG-7 and Delaunay^{View} use different approaches to representing semantic descriptions: MPEG-7 uses XML, while Delaunay^{View} uses RDF. An XML document is structured according to the tree paradigm: each element is a node and its children are the nodes that represent its subelements. An RDF document is structured according to the directed graph paradigm: each resource is a node and each

Figure 1. MediaLocator description scheme

```
<MediaLocator>
    <MediaURL>http://www.cs.uic.edu/~advis/ex.mpg</MediaURL>
    <MediaTime>
        <RelTime>PT12S</RelTime>
        <Duration>PT34S</Duration>
    </MediaTime>
</MediaLocator>
```

property is a labeled directed edge from the subject to the object of the RDF statement. Unlike XML, where schema and documents are separate trees, an RDF document and its schema can be thought of as a single connected graph. This property of RDF enables straightforward implementation of more powerful keyword searches as a means of selecting multimedia for presentation. Thus using RDF as an underlying description format gives users more flexibility in selecting content for presentation.

Another distinctive feature between MPEG-7 and Delaunay^{View} is the focus of the latter on multimedia presentation. A reference model for intelligent multimedia presentation systems encompasses an architecture consisting of control, content, design, realization, and presentation display layers (Bordegoni et al., 1997). The user interacts with the control layer to direct the process of generating the presentation. The content layer includes the content selection component that retrieves the content, the media allocation component that determines in what form content will be presented, and ordering components. The design layer produces the presentation layout and further defines how individual multimedia objects will be displayed. The realization layer produces the presentation from the layout information provided by the design layer. The presentation display layer displays the presentation. Individual layers interact with a knowledge server that maintains information about customization.

LayLab demonstrates an approach to multimedia presentation that makes use of constraint solving (Graf, 1995). This approach is based on primitive graphical constraints such as "under" or "beside" that can be aggregated into complex visual techniques (e.g., alignment, ordering, grouping, and balance). Constraint hierarchies can be defined to specify design alternatives and to resolve overconstrained states. Geometrical placement heuristics are constructs that combine constraints with control knowledge.

Additional work in multimedia presentation and information visualization can be found in Baral et al. (1998), Bes et al. (2001), Cruz and Lucas (1997), Pattison and Phillips (2001), Ram et al. (1999), Roth et al. (1996), Shih and Davis (1997), and Weitzman and Wittenburg (1994).

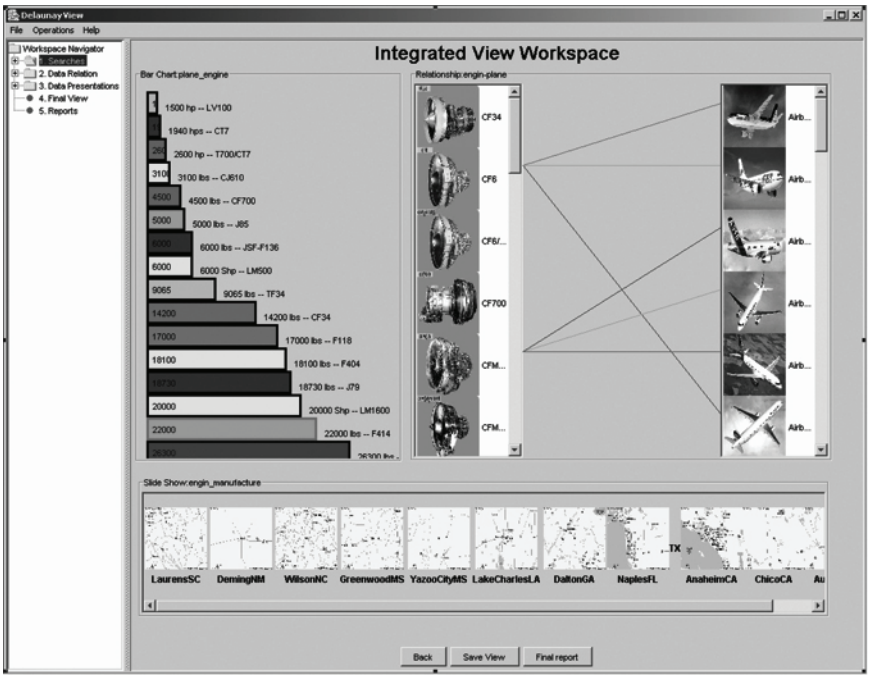
A PRAGMATIC APPROACH TO MULTIMEDIA PRESENTATION

In our approach to the design of a multimedia presentation system, we address the following challenges. Multimedia content is resident in distributed, heterogeneous, and autonomous sources. However, it is often necessary to access content from multiple sources. The data models and the design of the sources vary widely and are decided upon autonomously by the various entities that maintain them. Our approach accommodates this diversity by using RDFS to describe the multimedia sources in a simple and flexible way. The schemata are integrated into a single *global* schema that enables users to access the distributed and autonomous multimedia sources as if they were a single source. Another challenge is that the large volume of multimedia objects presented to the user makes it difficult to perceive and understand the relationships among them. Our system gives users the ability to construct customized layouts, thus making the semantic relationships among multimedia objects more obvious.

Case Study

This case study illustrates how multimedia can be retrieved and presented in an integrated view workspace using as example of a bill of materials for the aircraft industry. A bill of materials is a list of parts or components required to build a product. In Figure 2, the manufacturing of commercial airplanes is being planned using a coordinated visualization composed of three views:

Figure 2. A coordinated integrated visualization



a *bipartite graph*, a *bar chart*, and a *slide sorter*. The *bipartite graph* illustrates the part-subpart relationship between commercial aircrafts and their engines, the *bar chart* displays the number of engines currently available in the inventory of a plant or plants, and the *slide sorter* shows the maps associated with the manufacturing plants.

First, the user constructs a keyword query using the Search Workspace to obtain a data set. This process may be repeated several times to get data sets related to airplanes, engines, and plants. The user can preview the data retrieved from the query, further refine the query, and name the data set for future use.

Then, relationships are selected (if previously defined) or defined among the data sets, using metadata, a query, or user annotations. In the first two cases, the user selects a relationship that was provided by the integration layer. An example of such a relationship would be the connection that is established between the attribute engine of the airplane data set (containing one engine used in that airplane) and the engine data set. Other more complex relationships can be established using an RQL query.

Yet another type of relationship can be a connection that is established by the user. This interface is shown in Figure 3. In this figure and those that follow, the left panel contains the overall navigation mechanism associated with the interface, allowing for any other step of the querying or visualization process to be undertaken. Note that we chose the bipartite component to provide visual feedback when defining binary relationships. This is the same component that is used for the display of *bipartite graphs*.

The next step involves creating the views, which are built using templates. A data set can be applied to different templates to form different views. The interface of Figure 4 illustrates a *slide sorter* of the maps where the manufacturers of aircraft engines are located. In this process, data attributes of the data set are bound to visual attributes of the visual template. For example, the

Figure 3. Relation workspace

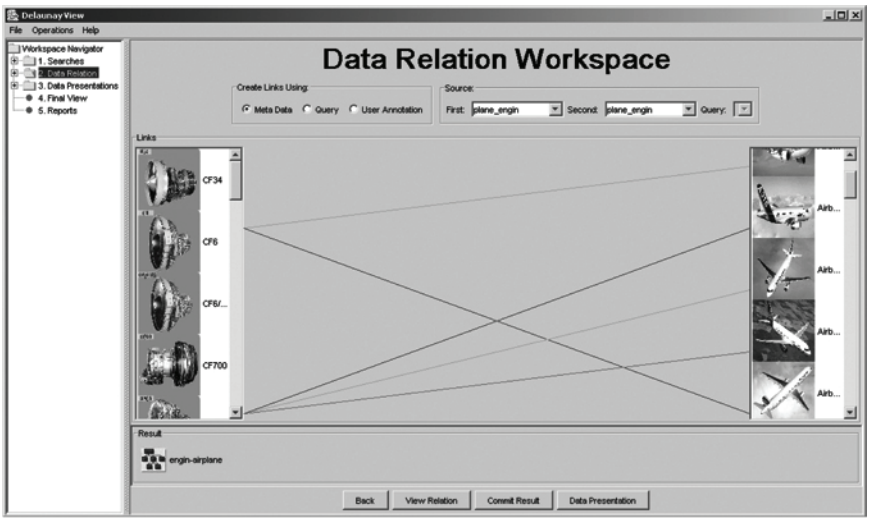
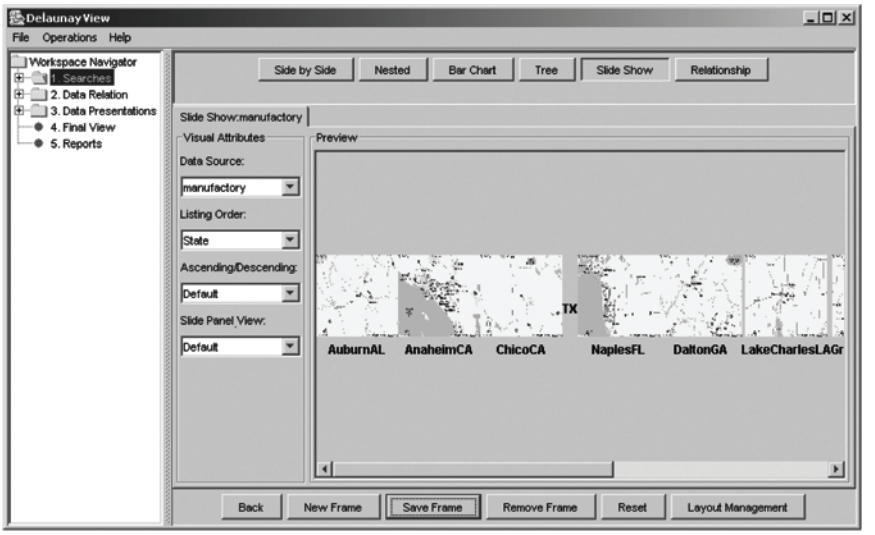


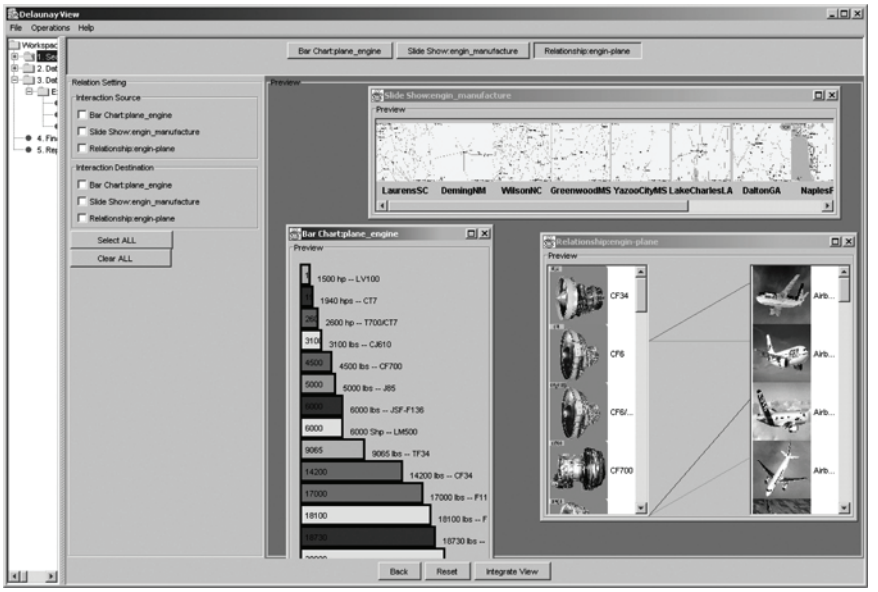
Figure 4. Construction of a view



passenger capacity of a plane can be applied to the height of a *bar chart*. The users also can further change the view to conform to their preferences, for example, by changing the orientation of a *bar chart* from vertical to horizontal. The sorter allows the thumbnails to be sorted by the values of any of the attributes of the objects that are depicted by the thumbnails. Individual views can be laid out anywhere on the panel as shown in Figure 5. The user selects the kind of dynamic interaction between every pair of views by using a simple customization panel.

In the integrated view, the coordination between individual views has been established. By selecting a manufacturing plant in the *slide sorter*, the bar displays the inventory situation of the selected plant; for example, the availability of each type of airplane engine. By selecting more

Figure 5. View layout



plants in the sorter, the *bar chart* can display the aggregate number of available engines over several plants for each type of airplane engine.

There are two ways of displaying relationships: they can be either represented within the same visualization (as in the *bipartite graph* of Figure 2) or as a dynamic relationship between two different views, as in the interaction between the bar chart and sorter views. Other interactions are possible in our case study. For example, the *bipartite graph* can also react to the user selections on the sorter. As more selections of plants are performed on the sorter, different types of engines produced by the selected manufacturer(s) appear highlighted. Moreover, the *bipartite graph* view can be refreshed to display only the relationship between the corresponding selected items in the two data sets.

System Architecture

The Delaunay^{View} system is composed of the *presentation*, *integration*, and *data* layers. The data layer consists of a number of autonomous and heterogeneous multimedia data sources that contain images and metadata. The integration layer connects the individual sources into a single *integrated virtual source* that makes multimedia from the distributed sources available to the presentation layer. The presentation layer includes user interface components that allow for users to query the multimedia sources and to specify how the images returned by the queries should be displayed and what should be the interaction among those images.

Data Layer

The data layer is comprised of a number of autonomous *multimedia sources* that contain images annotated with metadata. An image has a number of attributes associated with it. First there are the low-level features that can be extracted automatically. In addition, there are the application-dependent attributes that may include timestamps, provenance, text annotations, and any number of other relevant characteristics. All of these attributes determine the semantics of the image in the application context. Image attributes are described by an RDF schema. When an image is

Figure 6. Table arctic

imageId	source	image
QZ297492	23	X'22fa9c4 . . .'
S92305UN	11	X'ff0113a . . .'
7811300H	5	X'00f3439 . . .'
⋮	⋮	⋮

Figure 7. Arctic aerial photo metadata document

```
1   <?xml version="1.0" encoding="UTF-16"?>
2   <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3                xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
4                xmlns:aerial="http://www.cs.uic.edu/~advis/aerial.rdfs#">
5       <aerial:imageLocation>
6           <aerial:key>arctic.imageId</aerial.key>
7           <aerial:value>arctic.imageValue</aerial.value>
8       </aerial:imageLocation>
9       <aerial:image>
10          <aerial:timestamp>08/22/2004 14:07:23</aerial:timestamp>
11          <aerial:longitude>81°48'N </aerial:longitude>
12          <aerial:latitude>1°40'E </aerial:latitude>
13          <aerial:reference>QZ297492</aerial:reference>
14      </aerial:image>
15  </rdf:RDF>
```

added to a multimedia source, it is given a unique identifier and stored in a binary string format. A document fragment containing image metadata is created and stored in the database.

Example 1: Let us consider a multimedia source that contains aerial photos of the Arctic ice sheet used in climate research. The relevant metadata attributes include date and time a photo was taken and the latitude and longitude of the location where it was taken. A photo taken on "08/22/2003 14:07:23" at 81°48'N, 1°40'E is represented in the following way: the image file is stored in table arctic (Figure 6) with identifier "QZ297492". The RDF document fragment containing the metadata and referencing the image is shown in Figure 7.

In Figure 7, Line 3 declares namespace aerial which contains the source schema. Lines 5-8 contain the RDF fragment that describes the object-relational schema of table arctic; arctic. imageId contains a unique identifier and arctic. imageValue contains the image itself. Note that only the attributes that are a part of the reference to the image are described; that is, arctic.source is omitted. Lines 9-14 describe the metadata attributes of an aerial photo. They include timestamp, longitude, and latitude. The property reference does not describe a metadata attribute, but rather acts as a reference to the image object.

The source schema is shown in Figure 8. Lines 4-12 define class imageLocation with properties key and value. Lines 13-29 define class image

with properties timestamp, longitude, latitude and reference. Every time an image is added to the source an RDF fragment conforming to this schema is created and stored in the database. Although each fragment will contain a description of table arctic, this description will be stored only once.

We use the RDFSuite (Alexaki et al., 2000) to store RDF and RDFS data. The RDFSuite provides both persistent storage for the RDF and RDFS data and the implementation of the RQL language. The RDFSuite translates RDF data and schemata into object-relational format and stores them in a PostgreSQL database. The RQL interpreter generates SQL queries over the object-relational representation of RDF data and schema and processes RQL path expressions.

Integration Layer

The integration layer combines all multimedia sources into a single *integrated virtual source*. In the context of this layer, a multimedia source is a *local source* and its source schema is a *local schema*. The integrated virtual source is described by the *global schema*, which is obtained as a result

Figure 8. RDFS schema describing arctic photo metadata

```
 1  <?xml version="1.0" encoding="UTF-16"?>
 2  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 3           xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
 4      <rdfs:Class rdf:ID="imageLocation"/>
 5      <rdf:Property rdf:ID="key">
 6          <rdfs:domain rdf:resource="#imageLocation"/>
 7          <rdfs:range rdf:resource=
                "http://www.w3.org/1999/XMLSchema-datatypes#string"/>
 8      </rdf:Property>
 9      <rdf:Property rdf:ID="value">
10          <rdfs:domain rdf:resource="#imageLocation"/>
11          <rdfs:range rdf:resource=
                "http://www.w3.org/1999/XMLSchema-datatypes#string"/>
12      </rdf:Property>
13      <rdfs:Class rdf:ID="image"/>
14      <rdf:Property rdf:ID="timestamp">
15          <rdfs:domain rdf:resource="#image"/>
16          <rdfs:range rdf:resource=
                "http://www.w3.org/1999/XMLSchema-datatypes#string"/>
17      </rdf:Property>
18      <rdf:Property rdf:ID="longitude">
19          <rdfs:domain rdf:resource="#image"/>
20          <rdfs:range rdf:resource=
                "http://www.w3.org/1999/XMLSchema-datatypes#string"/>
21      </rdf:Property>
22      <rdf:Property rdf:ID="latitude">
23          <rdfs:domain rdf:resource="#image"/>
24          <rdfs:range rdf:resource=
                "http://www.w3.org/1999/XMLSchema-datatypes#string"/>
25      </rdf:Property>
26      <rdf:Property rdf:ID="reference">
27          <rdfs:domain rdf:resource="#image"/>
28          <rdfs:range rdf:resource=
                "http://www.w3.org/1999/XMLSchema-datatypes#string"/>
29      </rdf:Property>
30  </rdf:RDF>
```

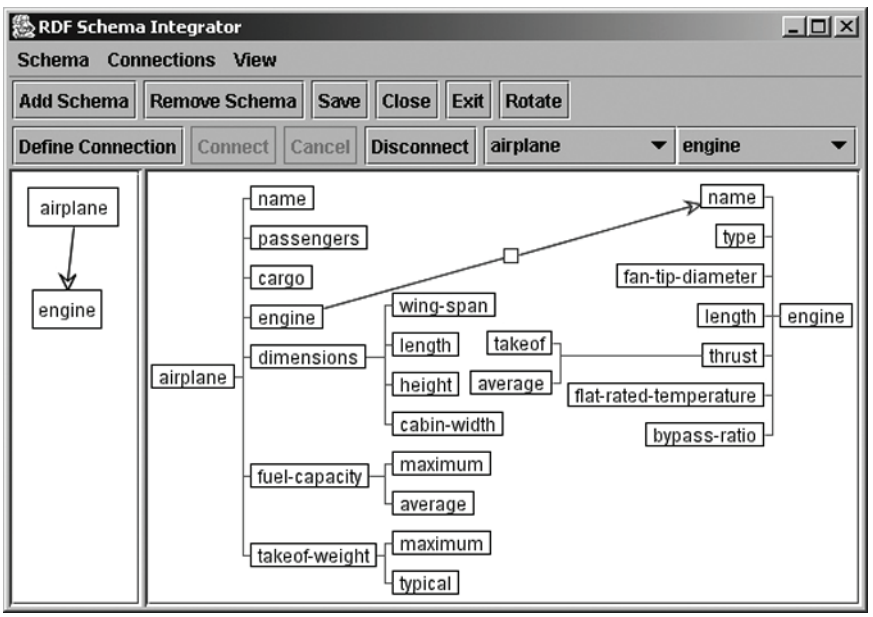
of the integration of the sources. Delaunay^{View} uses foreign key relationships to connect individual sources into the integrated virtual source. Implicit foreign key relationships exist between local sources, but they only become apparent when all local sources are considered as a whole. The global schema is built by explicitly defining foreign key relationships. A sequence of foreign key definitions yields a graph where the local schemata are the subgraphs and the foreign key relationships are the edges that connect them.

The foreign key relationships are defined with the help of the graphical integration tool of Figure 9. This tool provides a simple graphical representation of the schemata that are present in the system and enables the user to specify foreign key relationships between them. When the user imports a source into the system, its schema is represented on the left-hand side panel as a box. Individual schemata are displayed on the right-hand side pane as trees. The user defines a foreign key by selecting a node in each schema that participates in the relationship, and connecting them by an edge. Figure 9 shows how a foreign key relationship is defined between airplane and engine schemata. The edge between engine and name represents that relationship. The graphical integration tool generates an RDF document that describes all the foreign key relationships defined by the user.

The integration layer contains the *mediator engine* and the *schema repository.* The mediator engine receives queries from the presentation layer, issues queries to the data sources, and passes results back to the presentation layer. The schema repository contains the description of the global schema and the mappings from global to local schemata. The mediator engine receives queries in terms of the global schema, *global queries,* and translates them into queries in terms of the local schemata of the individual sources, *local queries*, using the information available from the schema repository. We demonstrate how local queries are obtained by the following example.

Example 2: The engine database and the airplane database are two local sources and engine name connects the local schemata. In the airplane

Figure 9. Graphical integration tool



schema (Figure 10), engine name is a foreign key and is represented by the property power-plant and in the engine schema (Figure 11) it is the key and is represented by the property name. Mappings from the global schema (Figure 12) to the local schema have the form ([global name], ([local name], [local schema])). We say that a class or a property in the global schema, *x*, maps to a local schema *S* when *(x, (y, S))* is in the set of the mappings. For this example, this set is:

(airplane, (airplane, S_1)),
(type, (type, S_1)),
(power-plant, (power-plant, S_1)),
(power-plant, (name, S_2)),
(engine, (engine, S_2)),
(thrust, (thrust, S_2)),
(name, (name, S_2))

All the mappings are one-to-one, except for the power-plant property that connects the two schemata; power-plant belongs to a set of foreign key constraints maintained by the schema repository. These constraints are used to connect the set of results from the local queries.

Figure 10. Airplane schema S1

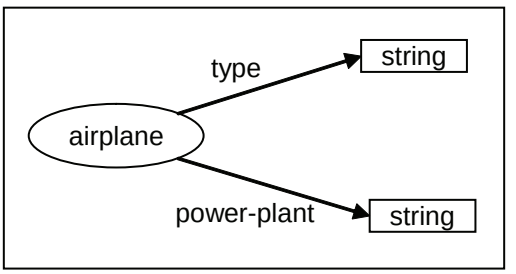


Figure 11. Engine schema S2

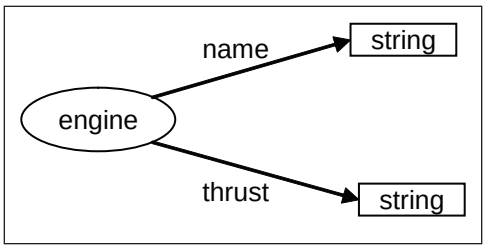
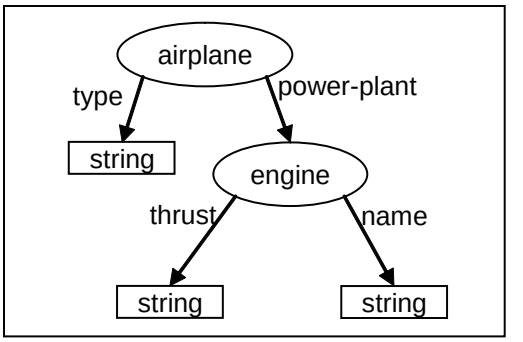


Figure 12. Global schema



The global query Q^G returns the types of airplanes that have engines with thrust of 115,000 lb:

```
select B
from {A}type{B}, {A}power-plant{C},
{C}thrust{D}
where D = "115000 lbs"
```

The mediator engine translates Q^G into Q^L_1, which is a query over the local schema S_1, and Q^L_2, which is a query over the local schema S_2. The *from clause* of Q^G contains three path expressions: {A}type{B}, which contains property type that maps to S_1 , {A}power-plant{C}, which contains property power-plant that maps both to S_1 and to S_2 , and {C}thrust{D}, which contains property thrust that maps to S_2 .

To obtain the *from clause* of a local query, the mediator engine selects those path expressions that contain classes or properties that map to the local schema. The *from clause* of Q^L_1 is: {A}type{B}, {A}power-plant{C}. Similarly, the *where clause* of a local query contains only those variables of the global *where clause* that appear in the local *from clause*. *D*, which is the only variable in the global *where clause*, does not appear in the *from clause* of Q^L_1, so the *where clause* of Q^L_1 is absent.

The *select clause* of a local query includes variables that appear in the global *select clause* and in the local *from clause*. *B* is a part of the global *select clause* and it appears in the *from clause* of Q^L_1, so it will appear in the *select clause* as well. In addition to the variables from the global *select clause*, a local *select clause* contains variables that are necessary to perform a join of the local results in order to obtain the global result. These are the variables is the local *from clause* that refer to elements of the foreign key constraint set. *C* refers to the value of power-plant, which is the only foreign key constraint, so *C* is included in the local *select clause*. Therefore, Q^L_1 is as follows:

```
select B, C
from {A}type{B}, {A}power-plant{C}
```

The *from clause* of Q^L_2 should include {A}power-plant{C} and {C}thrust{D}; power-plant maps to name and thrust maps to thrust in S_2. Since *D* in the global *where clause* maps to S_2, the local *where clause* contains *D* and the associated constraint: *D* = "115000 lbs". The global *select clause* does not contain any variables that map to S_2, so the local *select clause* contains only the foreign key constraint variable *C*. The intermediate version of Q^L_2 is:

```
select C
from {C}thrust{D}, {A}power-plant{C}
where D = "115000 lbs"
```

The intermediate version of Q^L_2 contains {A}power-plant{C} because power-plant maps to name in S_2. However, {A}power-plant{C} is a special case because it is a foreign key constraint: we must check whether variables A and C refer to resources that map to S_2. A refers to airplane, therefore it does not map to S_2 and {A}power-plant{C} should be removed from Q^L_2. The final version of Q^L_2 is:

```
select C
from {C}thrust{D}
where D = "115000 lbs"
```

In summary, the integration layer connects the local sources into the integrated virtual source and makes it available to the presentation layer. The interface between the integration and presentation layers includes the global schema provided by the integration layer, the queries issued by the presentation layer, and the results returned by the integration layer.

Presentation Layer

The *presentation layer* enables the user to query the distributed multimedia sources and to create complex multicomponent coordinated layouts to display the query results. The presentation layer sends user queries to the integration layer and receives the *data sets*, which are the query results. A *view* is created when a data set is attached to a presentation template that determines how the images in the data set are to be displayed. The user specifies the position and the orientation of the view and the dynamic interaction properties of views in the integrated layout.

Images and metadata are retrieved from the multimedia sources by means of RQL queries to the RDF multimedia annotations stored at the local sources. In addition to RQL queries, the user may issue keyword searches. A keyword search has three components: the keyword, the criteria, and the source. Any of the components is optional. A keyword will match the class or property names in the schema. The criteria match the values of properties in the metadata RDF document. The source restricts the results of the query to that multimedia source. The data sets returned by the integration layer are encapsulated in the *data descriptors* that associate the query, layout, and view coordination information with the data set. The following example illustrates how a keyword query gets translated into an RQL query:

Figure 13. Translation of a keyword search to an RQL query

```
1    SELECT DISTINCT kwResource
2    FROM {kwResource}^@kwProperty{kwValue},
3     {$kwClass}@kwProperty{$rangeClass}
4    WHERE ($kwClass LIKE "*airplane*" OR @kwProperty LIKE "*airplane*")
5            AND kwValue LIKE "*Boeing"
```

Example 3: The keyword search where keyword = "airplane", criteria = "Boeing", and source = "aircraftDataSource" returns resources that are of class "airplane" or have a property "airplane," have the property value "Boeing," and are located in the source "aircraftDataSource". This search is translated into the RQL query of Figure 13 and sent to source aircraftDataSource by the integration layer.

Delaunay^{View} includes predefined presentation templates that allow the user to build customized views. The user chooses attributes of the data set that correspond to template visual attributes. For example, a view can be defined by attaching the arctic photo dataset (see Example 1) to the *slide sorter* template, setting the *order-by* property of the view to the timestamp attribute of the data set, and setting the *image source* property of the view to the reference attribute of the data set. When a tuple in the data set is to be displayed, image references embedded in it are resolved and images are retrieved from multimedia sources.

The user may further customize views by specifying their orientation, position relative to each other, and coordination behavior. Views are coordinated by specifying a relationship between the *initiating view* and the *destination view.* The initiating view notifies the destination view of *initiating events.* An initiating event is the change of view state caused by a user action; selecting an image in a slide sorter, for example.

The destination view responds to initiation events by changing its own state according to the *reaction model* selected by the user. Each template defines a set of initiating events and reaction models.

In summary, semantics play a central role in Delaunay^{View} architecture. The data layer makes semantics available as the metadata descriptions and the local schemata. The integration layer enables the user to define the global schema that adds to the semantics provided by the data layer. The presentation layer uses semantics provided by the data and integration layers for source querying, view definition, and view coordination.

FUTURE WORK

Our future work will further address the decentralized nature of the data layer. Delaunay^{View} can be viewed as a single node in a network of multimedia sources. This network can be considered from two different points of view. From a centralized perspective, the goal is to create a single consistent global schema with which queries can be issued to the entire network as if it were formed by a single database. From a decentralized data acquisition point of view, the goal is to answer a query submitted at one of the nodes. The network becomes relevant when data are required that are not present at the local node.

In the centralized approach, knowledge of the entire global schema is required to answer a query while in the decentralized approach only the knowledge of paths to the required information is necessary. In the centralized approach, the global schema is static. Local sources are connected to each other one by one, resulting in the global schema that must be modified when a local schema is changed. Under the decentralized approach, the integration process can be performed at the

time the query is created (automatically in an ideal system) by discovering the data available at the other nodes. A centralized global schema must resolve inconsistencies in schema and data in a globally optimal manner. Under the decentralized approach inconsistencies have to be resolved only at the level of that node.

The goal of our future work will be to extend Delaunay^{View} to a decentralized peer-to-peer network. Under this architecture, the schema repository will connect to its neighbors to provide schema information to the mediator engine. Conceptually, a request for schema information will be recursively transmitted throughout the network to retrieve the current state of the distributed global schema, but our implementation will adapt optimization techniques from the peer-to-peer community to make schema retrieval efficient. The implementation of the mediator engine and the graphical integration tool will be modified to accommodate the new architecture.

Another goal is to incorporate MPEG-7 *Feature Extraction Tools* into the framework. Feature extraction can be incorporated into the implementation of the graphical integration tool to perform automatic feature extraction on the content of the new sources as they are added to the system. This capability will add another layer of metadata information that will enable users to search for content by specifying low-level features.

CONCLUSIONS

We have discussed our approach to multimedia presentation and querying from a semantic point of view, as implemented by our Delaunay^{View} system. Our paper describes how multimedia semantics can be used to enable access to distributed multimedia sources and to facilitate construction of coordinated views. Semantics are derived from the metadata descriptions of multimedia objects in the data layer. In the integration layer, schemata that describe the metadata are integrated into a single global schema that enables users to view a set of distributed multimedia sources as a single unified source.

In the presentation layer, the system provides a framework for creating customizable integrated layouts that highlight semantic relationships between the multimedia objects. The user can retrieve multimedia data sets by issuing RQL queries or keyword searches. The datasets thus obtained are mapped to presentation templates to create views. The position, the orientation, and the dynamic interaction of views can be interactively specified by the user. The view definition process involves the mapping of metadata attributes to the graphical attributes of a template. The view coordination process involves the association of metadata attributes from two datasets and the specification of how the corresponding views interact. By using the metadata attributes, both the view definition and the view coordination processes take advantage of the multimedia semantics.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation under Awards ITR-0326284 and EIA-0091489.

We are grateful to Yuan Feng Huang and to Vinay Bhat for their help in implementing the system, and to Sofia Alexaki, Vassilis Christophides, and Gregory Karvounarakis from the University of Crete for providing timely technical support of the RDFSuite.

REFERENCES

Alexaki, S., Christophides, V., Karvounarakis, G., Plexousakis, D., & Tolle, K. (2000). *The RDFSuite: Managing voluminous RDF description bases.* Technical report, Institute of Computer Science,

FORTH, Heraklion, Greece. Online at *http://www.ics.forth.gr/proj/isst/ RDF/RSSDB/rdfsuite.pdf*

Baral, C., Gonzalez, G., & Son, T. C. (1998). Design and implementation of display specifications for multimedia answers. In *Proceedings of the 14th International Conference on Data Engineering,* (pp. 558-565). IEEE Computer Society.

Bes, F., Jourdan, M., & Khantache, F. A. (2001) Generic architecture for automated construction of multimedia presentations. In the *Eighth International Conference on Multimedia Modeling.*

Bordegoni, M., Faconti, G., Feiner, S., Maybury, M., Rist, T., Ruggieri, S., et al. (1997). A standard reference model for intelligent multimedia presentation systems. *Computer Standards and Interfaces, 18*(6-7), 477-496.

Bray, T., Paoli, J., Sperberg-McQueen, C., & Maler, E. (2000). *Extensible markup language (XML) 1.0 (second edition).* W3C Recommendation 6 October 2000. Online at *http://www.w3.org/TR/2000/REC-xml-20001006*

Brickley, D., & Guha, R. (2001). *RDF vocabulary description language 1.0: RDF schema*. W3C Recommendation 10 February 2004. Online at *http://www.w3.org/TR/2004/REC-rdf-schema-20040210*

Cruz, I. F., & Huang, Y. F. (2004). A layered architecture for the exploration of heterogeneous information using coordinated views. In *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing* (to appear).

Cruz, I. F., & James, K. M. (1999). User interface for distributed multimedia database querying with mediator supported refinement. In *International Database Engineering and Application Symposium* (pp. 433-441).

Cruz, I. F., & Leveille, P. S. (2000). Implementation of a constraint-based visualization system. In *IEEE Symposium on Visual Languages* (pp. 13-20).

Cruz, I. F., & Lucas, W. T. (1997). A visual approach to multimedia querying and presentation. In *Proceedings of the Fifth ACM international conference on Multimedia* (pp. 109-120).

Fallside, D. (2001). *XML schema part 0: Primer.* W3C Recommendation, 2 May 2001. Online at *http://www.w3.org/TR/2001/REC-xmlschema-0-20010502*

Graf, W. H. (1995). The constraint-based layout framework LayLab and its applications. In *Proceedings of ACM Workshop on Effective Abstractions in Multimedia, Layout and Interaction,* San Francisco.

Hunter, J. (2001) An overview of the MPEG-7 Description definition language (DDL). *IEEE Transactions on Circuits and Systems for Video Technology, 11*(6), 765-772.

Karvounarakis, G., Alexaki, S., Christophides, V., Plexousakis, D., & Scholl, M. (2002). RQL: A declarative query language for RDF. In the *11th International World Wide Web Conference (WWW2002).*

Klyne, G., & Carroll, J. (2004). *Resource description framework (RDF): Concepts and abstract syntax.* W3C Recommendation 10 February 2004. Online at *http://www.w3.org/TR/2004/REC-rdf-concepts-20040210*

Martínez, J. M. (Ed.) (2003) *MPEG-7 overview.* ISO/IEC JTC1/SC29/WG11N5525.

Pattison, T., & Phillips, M. (2001) View coordination architecture for information visualisation. In *Australian Symposium on Information Visualisation, 9,* 165-169.

Ram, A., Catrambone, R., Guzdial, M.J., Kehoe, C.M., McCrickard, D.S., & Stasko, J. T. (1999). PML: Adding flexibility to multimedia presentations. *IEEE Multimedia, 6*(2), 40-52.

Roth, S. F., Lucas, P., Senn, J. A., Gomberg, C. C., Burks, M. B., Stroffolino, P. J., et al. (1996)

Visage: A user interface environment for exploring information. In *Information Visualization*, 3-12.

Salembier, P., & Smith, J. R. (2001). MPEG-7 multimedia description Schemes. *IEEE Transactions on Circuits and Systems for Video Technology, 11*(6), 748-759.

Shih, T. K., & Davis, R. E. (1997). IMMPS: A multimedia presentation design system. *IEEE Multimedia, 4*(2), 67-78.

Weitzman, L., & Wittenburg, K. (1994). Automatic presentation of multimedia documents using relational grammars. In *Proceedings of the Second ACM International Conference on Multimedia* (pp. 443-451). ACM Press.

This work was previously published in Managing Multimedia Semantics, edited by U. Srinivasan and S. Nepal, pp. 333-350, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 4.6
OntoMedia:
Semantic Multimedia Metadata Integration and Organization

Bodo Hüsemann
Informationsfabrik GmbH, Münster, Germany

Gottfried Vossen
University of Münster, Germany,
University of Waikato, New Zealand

ABSTRACT

Digital multimedia devices for private usage have nowadays left their analogous counterparts behind. Our homes increasingly incorporate a digital communication infrastructure for interconnecting their various devices inside the house, and this infrastructure also gives access to external information via the Internet. Thus, we now store much of our personal information in digital form and information services like newspaper, radio, or TV are digitally available. The capacity of digital storage devices provides enormous space at affordable prices to save all this digital information into a giant personal multimedia archive; however, as is well known, the more data we store, the less information we have at our disposal. In this article, an approach to personal information management is described that is based on Semantic Web technology. In particular, the design decisions behind, and core features of OntoMedia, a software system based on a multimedia ontology that is intended for personal usage, are presented. OntoMedia supports users organizing personal multimedia archives using ontology powered metadata extraction and integration technology. The application includes a novel, easy to use graphical user interface to organize documents in categories and browse/query the semantic database.

INTRODUCTION

Digital multimedia devices for private usage, including digital cameras, camcorders, or mp3 players, have nowadays left their analogous counterparts behind. Our homes increasingly incorporate a digital communication infrastructure for interconnecting their various devices inside the house, and this infrastructure also gives access to external information via the Internet. Thus,

we now store much of our personal information in digital form and information services like newspaper, radio, or TV are digitally available. The capacity of digital storage devices provides enormous space at affordable prices to save all this digital information into a giant personal multimedia archive (cf. the MyLifeBits project, see Gemmell, Bell, & Lueder, 2006); however, as is well known, the more data we store, the less information we have at our disposal. In this article, we describe an approach to personal information management that is based on Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) technology. In particular, we present the design decisions behind and core features of *OntoMedia*, a software system based on a multimedia ontology that is intended for personal usage. *Personal information management* (PIM) refers to all activities related to the creation, storage, organisation, use, and retrieval of digital information in a personal usage-context. In this article, we focus digital information represented by multimedia documents like digital images, music, texts, or video recordings.

The creation and storage of digital content was never easier, but with increasing volume, the management of multimedia archives turns out to be a non-trivial task with vital research problems. Indeed, ACM's SIGMM ranks "to make capturing, storing, finding, and using digital media an everyday occurrence in our computing environment" one of three major goals for multimedia research for the future (Rowe & Jain, 2005).

With the software application OntoMedia, we address this vision to ease the management of multimedia archives for private users and personal usage (Hüsemann & Vossen, 2004). We have developed this application with two main goals in mind:

1. Automatic extraction and semantic integration of multimedia metadata should be supported.
2. Organisation and search of multimedia documents based on semantic metadata is a necessity.

With *multimedia metadata*, we denote every form of digital information about some multimedia document. The enabling technology to integrate, process, and query metadata for multimedia is delivered by current languages and tools of the Semantic Web.

In Figure 1, we exhibit the schematic evaluation of a user query searching for an instance of concept "person," where the OntoMedia Core Ontology is used to determine semantic relationships to similar concepts in multimedia metadata or the Semantic Web. Basically, this kind of semantic inference is used by OntoMedia to find files related to semantic queries given by the user.

The article is organised as follows: In the second section we briefly review related work. In the third section we describe the OntoMedia application from a user's point of view, and we also explain its system architecture. In the fourth section we study metadata generation and import, and in the fifth section we look at search processes within OntoMedia. With the sixth section, we discuss or experiences with OntoMedia and our approach in general. Finally, we draw some conclusions in the seventh section.

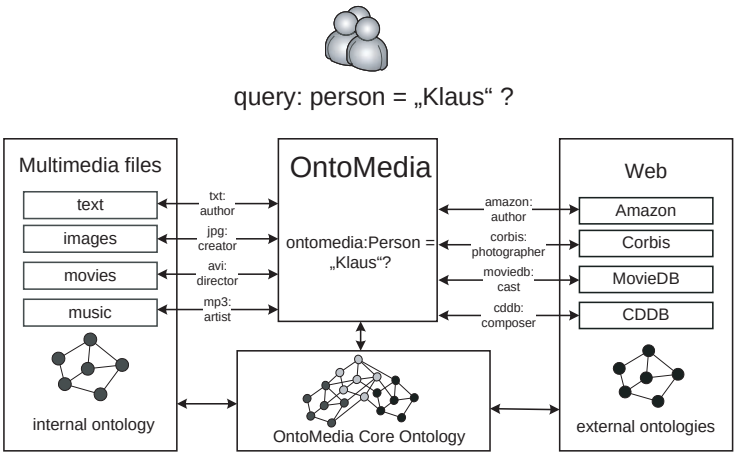
RELATED WORK

The extraction and integration of multimedia metadata has been a central goal for various information systems in the past up to the current hype for desktop indexing and search tools for private users. We summarise the functionality of those systems by the following core features:

- Metadata generation
- Document classification
- Document retrieval

Metadata generation (Handschuh & Staab, 2003) is done with automatic indexing techniques or by manual annotation. There are three different approaches to get metadata for multimedia

Figure 1. Semantic metadata integration with the OntoMedia Core Ontology



documents with software systems: the first approach is content driven where metadata is computed by content analysis mechanisms (e.g., natural language processing of audio data). This analysis produces, for example, text extracts for text documents (Appelt, Hobbs, Bear, Israel, & Tyson, 1993), colour histograms for images (Bimbo, 1999), or frequency histograms for music (Foote, 1997).

The second approach simply extracts already present metadata from a document which is included as part of the document container. This metadata is typically some sort of textual attribute value description for a document. The third is to query external information sources for metadata about a document. This ideally requires a unique identification technique for multimedia documents (e.g., some content dependent hash value, cf. Cano, Batlle, Kalker, & Haitsma, 2005) which can be used to query an external service to retrieve metadata for a document.

For huge personal multimedia archives, manual metadata annotation is very time consuming and content analysis for multimedia documents other than texts is rather ineffective. Thus, OntoMedia follows the second alternative, building a database by automatic metadata extraction. In this article, we will show how we use ontologies to define semantic relations between metadata attributes of different document types with concepts of the multimedia core ontology.

Document classification is a method to associate classes (= categories) with documents. All documents inside a class build a homogeneous set according to some classification criteria. With OntoMedia, we use an ontology-based classification method where documents are directly classified using their metadata. Additionally, the inherent semantics of the ontology provides indirect classifications, which the user can use to query the database. Thus, users can query the database using well-known concepts of the multimedia ontology. Additionally, users manually classify documents and personalise classification structures as they like.

Finally, document retrieval methods select relevant documents of a set of documents according to some retrieval criteria given, for example, by a user defined query. The novel document retrieval feature of OntoMedia is its graphical user interface that supports users to query the database using classes of the multimedia ontology. Users can select a collection of categories with a graphical representation of the ontology and define metadata filters based on properties to search for documents. Given that, OntoMedia

will search for related documents using the formal semantics of the multimedia ontology.

At present, users typically arrange their multimedia collections in file systems, which provide poor naming mechanisms and hierarchical directory structures for organisation and search. Although this approach looks sufficient at first sight there are many problems involved which complicate the management of large multimedia collections (Hüsemann & Vossen, 2004):

- The logical organisation is bound to the underlying physical storage system.
- Categorisation is limited to strict classification hierarchies based on directory structures.
- Document-inherent metadata (e.g., ID3 tags in MP3 files) remains unused and is only available to media-specific applications (e.g., an MP3 player).
- Identification based on file names alone is often not globally consistent (e.g., duplicates are possible).

To overcome these problems, desktop indexing tools try to help users with searching for files. In Table 1, we summarise the functionality provided by desktop indexing tools that are currently available and compare those with OntoMedia.

The tools shown in Table 1 do not support personalised categorisation structures to manu-

Table 1. Functionality of current desktop indexing tools

	Google Desktop http://desktop.google.com	Copernic Desktop http://www.copernic.com/en/products/desktop-search	MSN Desktop http://toolbar.msn.com/	x-friend http://www.x-friend.de	OntoMedia
Metadata generation					
Index engine	Google	Copernic	MS iFilter	Lucene	OntoMedia
Automatic indexing	+	+	+	+	+
Index format	proprietary	proprietary	proprietary	proprietary	RDF
Supported indexed formats (content/ metadata)					
Archive	-/+	-/+	-/-	-/+	-/-
Audio	-/+	-/+	-/+	-/+	-/+
Chat	+/+	-/-	-/-	-/-	-/-
Image	-/+	-/+	-/-	-/+	-/+
Mail	+/+	+/+	+/+	+/+	-/-
Text	+/+	+/+	+/+	+/+	-/+
Video	-/+	-/+	-/-	-/-	-/+
Other	Web, History, RSS	Web, History, Contacts	Web, Encarta, Contacts	Web, RSS	RDF
Classification					
Manual classification	-	-	-	-	+
Automatic classification	+ (Filetype)	+ (Filetype)	+ (Filetype)	+ (Filetype)	+ (Metadata)
Retrieval / Search					
Browse	-	+	-	+ (Web)	+
Boolean	+	+	+	+	+
Metadata	-	+	-	-	+
Ranked results	+ (without measure)	-	+ (measure included)	+ (without measure)	-

ally classify, browse, and search file collections. Along with the automatic classification features of OntoMedia, users can build up their own categorisation structures that can be browsed to categorise and search for files. Thus, users don't need additional software anymore to annotate (e.g., photo-archives by keywords). The categorisation graph of OntoMedia is independent of physical storage structures and not limited to simple hierarchies. Sub-categories can have more than one super-category, which allows for much more powerful organisation structures.

In contrast to the applications shown that use proprietary indexing formats, we use standardised languages of the Semantic Web, which allow for semantic integration of metadata. OntoMedia features a multimedia core ontology to describe all metadata attributes and their semantic connection to enable metadata search processes. The whole information is stored with non-proprietary Resource Description Format (RDF) (cf. Becket, 2004; Lassila & Swick, 2004), which can be exchanged with other Semantic Web enabled applications. Furthermore, the database includes search facilities which feature semantic reasoning capabilities to support the user while searching for relevant attributes inside the ontology. After indexing, OntoMedia initially classifies all documents by rebuilding the current directory structures in the category graph as basis.

ONTOMEDIA FROM A USER'S PERSPECTIVE

In this section, we will use the main user interface of OntoMedia for a guided tour to the functionality of the application. The main user interface of OntoMedia is divided into six different parts shown in Figure 2.

The main part of the user interface between status and menu bar is divided into four different areas where the left half shows schema information of the current ontology and the right half shows instance data about indexed documents in the database.

The left hand side shows the category graph (area 2) and the metadata properties (area 4) of the OntoMedia database. The category graph is a visualisation of the class hierarchy given by the multimedia ontology to categorise files with OntoMedia. The user can navigate, select, and modify the category graph in order to search or organise document collections. The different colours of categories denote different select modes we will explain in detail later. In area 4 (below the category graph), users can define metadata-based search criteria using the displayed metadata property tree. The user can navigate all metadata properties available in the database and define search criteria to filter the results of a search request.

On the right hand side, the user interface shows a tabular listing of files (the file browser) in the

Figure 2. Overview of the OntoMedia user interface

database (area 3), and a playback panel with further file specific details (area 5). The file listing contains all files relevant to the current selection of categories and defined filter criteria. The File Viewer area contains different panels to display all metadata about a selected file, display a set of images, or playback audio/video files selected in the File Browser/Basket.

The menu and toolbar at the top of the user interface (cf. area 1 in Figure 2) provides access to the main functionality concerning data import/export, organisation, and search commands available to the user. The user can export the database in RDF/XML format or import any RDF data to the current database. Thus it is possible to include any RDF based information which is useful (e.g., to import external databases to OntoMedia). The status bar contains information about the current size of the database, and the memory resources used by OntoMedia.

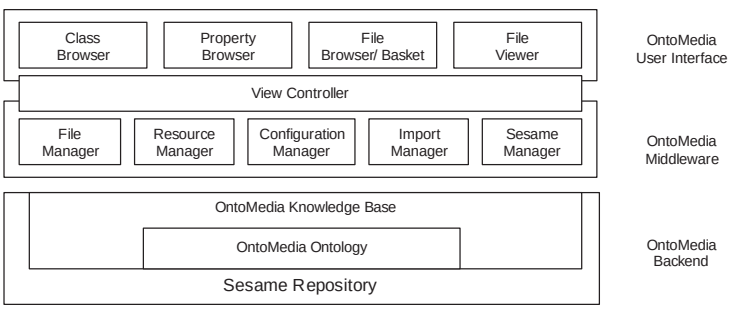
We next give an overview to the software architecture of OntoMedia and some of its external components. The main architecture of OntoMedia is broken down into three major layers that are *user interface*, *middleware*, and *backend*. In principle these parts, which are shown in Figure 3, build a classical three-tier architecture, where all parts could be physically distributed.

Backend: The backend layer performs the centralised data management of the application, which is realised by an external component: the Sesame RDF framework (Broekstra, Kampman, & Harmelen, 2002). The Sesame Framework provides a storage and inference abstraction layer for RDF data. The framework supports different physical data storage options, where users can choose to query and store RDF repositories with memory model stores, native file based stores, or with conventional object relational databases. The backend is completely implemented by Sesame, and can reside on a separate physical system in a multi user environment.

Middleware: The middleware layer of OntoMedia implements high level services of the application which include:

- **File management:** The File Manager provides physical abstraction of files and paths.
- **Configuration management:** the Configuration Manager gives persistent access to global parameters and variables of the application.
- **Resource management:** The Resource Manager is a background process which provides several runtime monitoring services (memory usage, database statistics) to the application.
- **Import management:** The Import Manager handles import processes of the application in a separate thread. It invokes the necessary extractor plugins to scan files for inherent

Figure 3. Overview of the OntoMedia software architecture



metadata, and delivers extracted metadata to the Sesame Manager.

- **Sesame management:** The Sesame Manager is the central access service of the middleware to the OntoMedia backend. The Sesame Manager is subdivided in:
  - Connection Manager to provide physical connection to the backend.
  - Update Manager to perform services with write access to the database.
  - Query Manager to deliver read services evaluating query request.

User Interface: The User Interface Layer (UIL) of OntoMedia consists of several components to browse/query the database and display information about files in the database with related metadata. The UIL is connected to the OntoMedia middleware with the View Controller. The View Controller is a central manager for all user interactions, which need complex process control and communication with the middleware. It delegates queries and changes requests to the backend and handles high level services requests of the UIL through invocation of the other middleware services available.

In detail, the UIL consists of the following main components:

- **Class Browser:** The Class Browser displays a graphical representation of the category hierarchy, and is based on the graphic framework Touchgraph (http://www.touchgraph.com/). The nodes of the graph represent categories, which are connected via directed edges, where edges point from sub categories to super categories. The graph is a non-strict poly hierarchy where sub categories potentially relate to several super categories but form no cycles.

  The Class Browser component uses a spring layout algorithm (Eades, 1984) to display the category graph, which maintains good readability with complex graphs. The layout strategy calculates graph animation based on a force model, where the edges between nodes form ideal rubber bands to drag them to an overall balanced state.

  In Figure 4, we see a simple category graph where the root of the hierarchy is denoted "***" (cf. node 1 in step A).

  The figure illustrates navigation through the animated graph in four subsequent steps A to D, where numbers show the order of category selections in time done by the user. As the user selects a category in the graph, the system expands all its sub categories. To focus the animation on categories of interest the system collapses all categories in the graph which are not part of selected path.

  The user can modify the category graph to fit personal needs, and thus build up own organisation structures which will be discussed in detail later.
- **Property Browser:** The Property Browser component shows the available metadata properties and their hierarchical relation in a tree representation. The property tree is a hierarchical representation of the property hierarchy of the OntoMedia Core Ontology (OMCO). The user can specify property filters by exploring the tree for relevant properties, and defining filter criteria using textual search patterns. In Figure 5, the user has specified a filter "*fire*" on the attribute "External Metadata," and browses the property "Artist" for available values in the database.
- **File Browser/ Basket:** The File Browser shows a tabular listing of files and their associated metadata in the database. The columns of the File Browser represent selected properties of the OMCO, which contain available information to every entry in the current list. The user can change the displayed columns of the table via drag & drop and is able to define property sets,

Figure 4. Navigation of the category hierarchy with the Class Browser

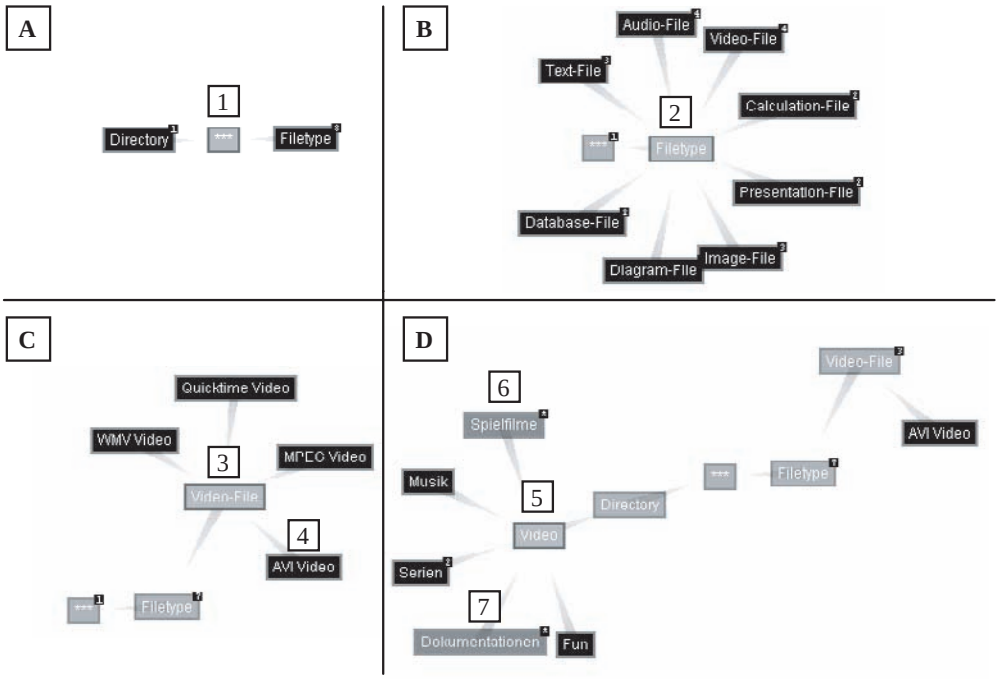
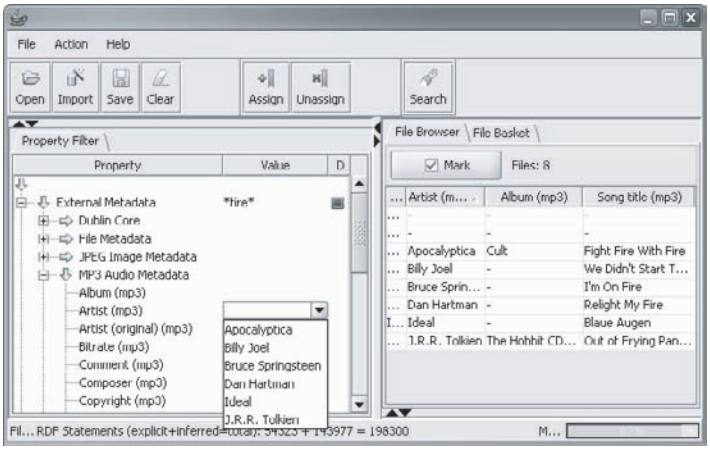


Figure 5. Definition of filter criteria with the Property Browser



which can be related to hotkeys for quick reference. To change the display of the listing, the user can sort the table by multiple columns or change their order. In addition is possible to group all entries in table by a choosing a grouping property column. The table shows multi valued properties by drop down lists inside the cells. The second file listing available (called File Basket) is a convenience control to help users building file collections with OntoMedia. Users can add selected files from the File Browser to the File Basket in order to iteratively build up file sets with multiple search processes.

- **File Viewer:** After selecting files with the File Browser users choose between dif-

ferent viewers to show more details about associated metadata and their contents. The following components are available:

- **Metadata Viewer:** The metadata viewer shows a list of all available metadata for a selected file. The list shows multi-valued properties with drop down boxes. This list is read only at the moment, but will provide modify access in future versions of OntoMedia.
- **Multimedia Viewer:** The multimedia viewer contains different controls to playback audio and video contents. OntoMedia supports playback of Quicktime, DivX, and MPEG video files within the application using the Java Media Framework (JMF) (http://java.sun.com/products/java media/jmf/index.jsp), and the QuickTime Java SDK (http://www.apple.com/quicktime/qtjava) in conjunction with 3vix decoders (http://www.3ivx.com/). As for audio contents OntoMedia makes use of the JLayer (http://www.javazoom.net/javalayer/javalayer.html) library to play MPEG 1 Layer 3 (MP3) contents. Other uncompressed formats (like WAVE) are handled by JMF. The multimedia viewer includes controls to manage playback queues, volume, and full screen playback.
- **Image viewer:** The image viewer provides simple browsing functionality of image collections in a customisable thumbnail view. The user can control the size of images in the list and switch to a full screen slideshow.

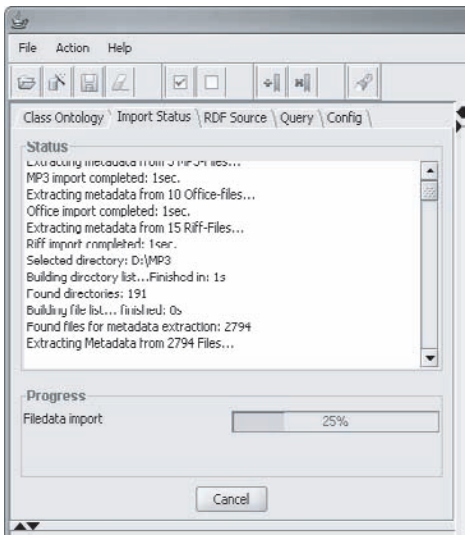
METADATA GENERATION AND IMPORT

Manual annotation of files with metadata is a very time consuming task for voluminous collections. Thus, OntoMedia provides automatic metadata extractors for many file formats, which scan files for included format specific metadata. Once files are scanned with OntoMedia, they can be found by searching the indexed metadata in the database using the OMCO.

The generation and import of metadata is triggered by the user via the tool bar and follows a simple sequential three-step process coordinated by the Import Manager:

1. **Definition of the import task:** The user chooses a directory for the scan process within a conventional file dialog. The system generates an import task for that directory.
2. **Task analysis and preparation:** The selected directory is scanned recursively for files including all inherent subdirectories.

Figure 6. Import status panel which reports progress of metadata import tasks



Every file in the generated file list is queued according to its filename extension to work lists for registered metadata extractors, which can handle files with that extension.

3. **Task processing:** The Import Manager invokes all responsible extractors to process their work list. Every metadata extractor scans all files in its work list for metadata and reports its progress to the import status panel (cf. Figure 6). The extracted metadata will be transformed and integrated according to the OMCO. In the end all extracted, transformed, and integrated metadata is delivered to the Sesame Manager and stored in the database.

The metadata included with files needs to be transformed by an extractor into RDF-based description triples and semantically integrated with the OMCO. For the semantic integration of the imported metadata, the current set of extractors use mappings from specific import ontologies to the OMCO.

In Figure 7, we see a sample of an RDFS import ontology defined by the MP3 metadata extractor used by OntoMedia. The example uses XML entities to encode URL references in XML attribute values where the entity &mp3; (cf. line 1) points to the namespace http://www.ontomedia.de/mp3 of the mp3 extractor and &omco; (cf. line 3) references the OMCO namespace http://www.ontomedia.de/omco. The ontology defines the property &mp3;#copyright which is mapped to the OMCO property &omco;#copyright through the specification of a RDFS subPropertyOf relation (cf. line 3). In addition to this semantic mapping rule the mp3 extractor defines an own application specific RDF property <conv:name> (cf. line 4) which contains a mapping to proprietary attribute name "Copyright" reported by the used mp3 extraction library API. Thus all necessary mappings are defined outside the application code using standardised ontology languages which can be easily processed by RDF aware standard software.

We mention that we have discussed the design of the ontology underlying OntoMedia in an earlier paper (Hüsemann & Vossen 2005).

INFORMATION SEARCH IN ONTOMEDIA

In this section, we will give an introduction to the organisation and search functionality build in with OntoMedia. There are three different types of main management activities, which users will typically invoke with OntoMedia:

1. **Management of categorisation structures:** Users can organise file collections in categories using the category browser. With the management functionality of the category browser, the user can adapt the category graph to their personal needs.

Figure 7. Definition of the RDF property "&mp3;#copyright" and its mapping to OMCO property "&omco;#copyright"

```
1.<rdf:Property about="&mp3;#copyright">
2. <rdfs:subPropertyOf rdf:resource="&mp3;#mp3Metadata"/>
3. <rdfs:subPropertyOf rdf:resource="&omco;#copyright"/>
4. <conv:name>Copyright</conv:name>
5. <rdfs:range rdf:resource="&xsd;#string"/>
6. <rdfs:domain rdf:resource="&omco;#MP3 Audio"/>
7. <rdfs:label xml:lang="de">Copyright (mp3)</rdfs:label>
8. <rdfs:label xml:lang="en">Copyright (mp3)</rdfs:label>
9.</rdf:Property>
```

2. **File categorisation:** Given a category graph, build with the category browser users can manually assign categories to files and thus building a powerful classification system.
3. **File search:** When users have organised their file collections with OntoMedia, they can access the powerful OntoMedia search facilities using the Category Browser or the Property Filter to search for documents.

We will now look at those core activities in more detail, and we will explain some related conceptual design decisions.

Management of Categorisation Structures

The OMCO ontology used by OntoMedia defines basic categories suitable as a generic starting point for building up personalised organisation structures within the category browser. The management of the category graph with OntoMedia is performed by the following action types: manipulation of categories and manipulation of category relations.

Manipulation of Categories

The first thing with building personal categorisation structures using OntoMedia is to add new categories to the graph. New categories can be inserted by selecting a super category with the mouse and dragging away the new category from the selected one (cf. Figure 8a). While accessing the context menu pointing to a selected category, the user can choose further manipulation actions including commands to rename or delete categories from the graph (cf. Figure 8b). If the user deletes a category from the graph, all categories in the related sub hierarchy will be deleted too.

Figure 8. Manipulation of categories adding new categories: (a) and renaming/deleting categories (b) via context menus

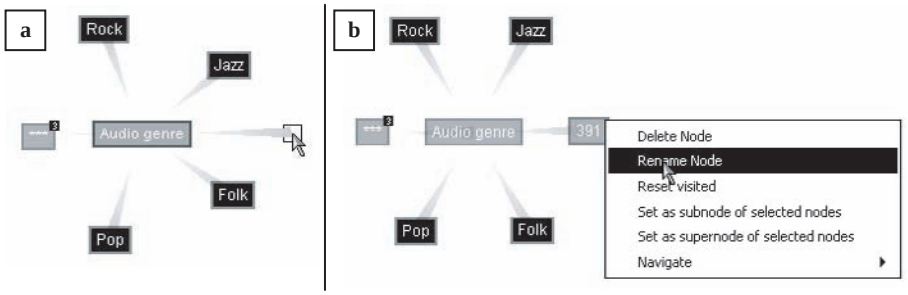
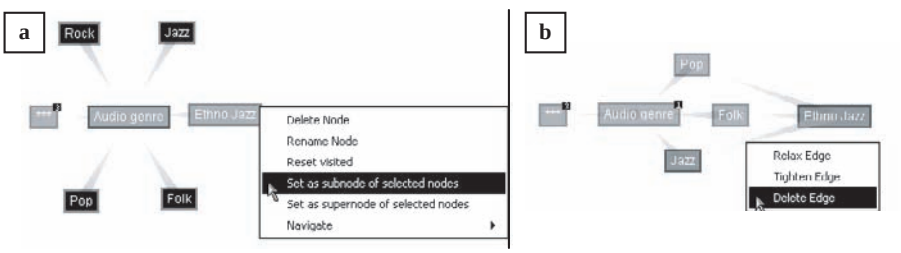


Figure 9. Manipulation of category relations using the context menu to add (a) or delete (b) category relations in the graph



Manipulation of Category Relations

Relationships between categories can be added (cf. Figure 9a) and deleted (cf. Figure 9b) using their context menus in the Category Browser. If the user detaches a sub category from its super category every file in every sub category of the related sub hierarchy will recursively loose its relation to the super category, too.

File Categorisation

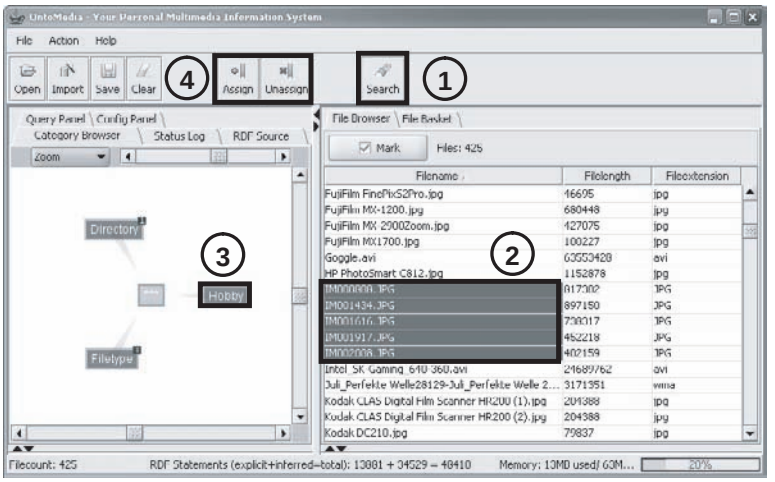
The (de-)categorisation of files with OntoMedia is a simple four step sequential process (cf. Figure 10):

1. **File search:** the user searches for potential relevant files using the file search capabilities of OntoMedia. The File Browser displays the list of all files relevant to the current search criteria defined with the Category Browser and the Property Filter.
2. **File selection:** Using the File Browser or the File Basket users select all files in the current list they want to process within the categorisation process.
3. **Category selection:** The user marks all relevant categories in the graph in order to define the target for the categorisation. All marked categories have red colour in the graph.
4. **Categorisation processing:** With the last step the user triggers the categorisation process by choosing the "Assign" or "Unassign" command in the toolbar.

OntoMedia supports searching in the OntoMedia database by using the Category Browser and the Property Filter to define search criteria. In Figure 11, we can see a simple sequential search process within OntoMedia with the following steps:

1. **Category restriction:** The user restricts the search to a selected set of categories.
2. **Property restriction:** The user specifies textual constraints with the property filter.
3. **Query processing:** The user triggers query evaluation via the search command button.
4. **Result display:** The user retrieves the result which is displayed in the File Browser.

Figure 10. File categorization with OntoMedia



File Search

The user can activate two different select modes in the category browser, which define necessary and optional category for files in the result of a query. All necessary categories are displayed red and all optional categories are displayed green in the graph. Yellow categories get indirectly selected by subcategories or lie on the path to the current active node in the graph. All other categories in the graph have blue colour. In addition to category selections the user can define, textual constrains using the Property Filter in OntoMedia. It is possible to use wildcards in a query indicated by an "*" as commonly used in keyword based retrieval systems.

In Figure 11, we can see the definition of the following sample query with OntoMedia:

Q = "Give me all files which are related to the categories "Audio File" and at least one of the categories "Rock" or "Folk" where the property "Album (mp3) contains the string "the best."

A generic representation of this query using a simple Boolean retrieval model where categorisation is modelled by the attribute "cat" would be as follows:

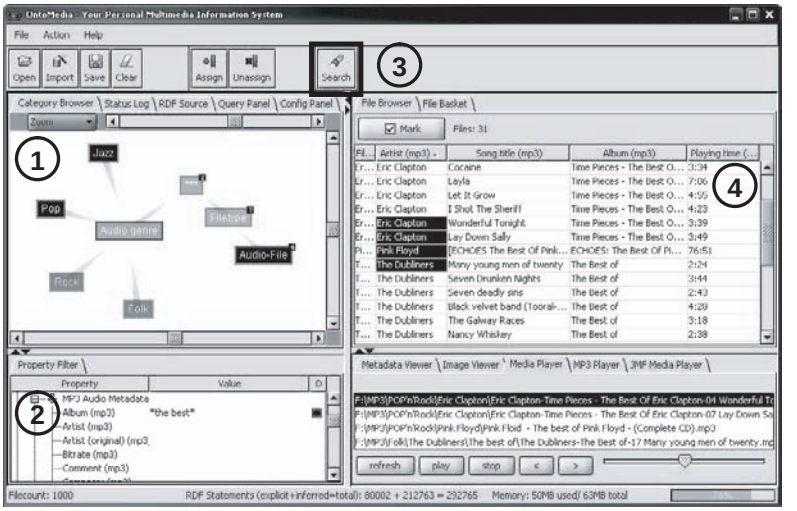
```
Q':= (cat, "Audio File") AND
     ((cat, "Rock") OR (cat, "Folk")) AND
     (Album (mp3), "*the best")).
```

The Sesame Framework uses common RDFS semantics to evaluate queries of OntoMedia. Using RDFS semantics inferencing services will expand user queries to all related sub categories and sub properties defined in a user query using the ontology of the database. If the user queries a super property `artist` the system will expand the query to all related sub properties of artist (e.g., `painter`, `composer`, `and photograph`) too. The same is true for categories where RDFS semantics relates all files in a sub category implicitly to all super categories.

DISCUSSION

OntoMedia is targeted at use case scenarios of private users. We have not done a systematic user evaluation yet, so we will share some general experiences we have gained from conferences and our own use next.

Figure 11. Example search process with OntoMedia



From a functional point, of view users like the category graph to define queries and enjoy its layout animation. Especially non-technical users appreciate the metadata property tree, as they can browse for relevant properties on higher abstraction levels without knowledge of the exact technical denotation typical for detailed metadata.

We consider the graphical user interface very useful to organise medium sized category graphs (< 200 categories, max. 8 levels). If a category has many subcategories (> 30) on the next level the radial layout has disadvantages as we didn't implemented sorting to simplify systematic search by enumeration.

From a technical point of view, we have tested the system with medium-size multimedia archives up to 10,000 files on standard PC equipment. The underlying RDF Framework Sesame performs brilliant at querying persistent memory stores (query evaluation takes less than one second). Nevertheless, the current Sesame implementation has significant performance problems with update operations. So changes of the categorisation graph like deleting categories or modifying document categorisations can be rather slow (update operations can take > 1 minute). Hopefully, the Sesame community will provide some solution to those update performance problems in upcoming releases.

Compared to standard desktop indexing tools, OntoMedia demonstrates the effective use of Semantic Web technology to improve the management of personal multimedia collections. Users can browse the available metadata by exploiting clearly defined semantic relations between available classes and attributes in the multimedia ontology. This approach is much more effective then simple unstructured text queries. The OntoMedia Core Ontology defines a standard vocabulary for the technical properties and content of multimedia. The concepts of the ontology are easy to understand for non-experienced users, which can thus query for documents with an easy to understand vocabulary without expert-knowledge of the metadata included in specific file-formats.

At the moment, OntoMedia targets the multimedia environment of private users while its general architecture and use of Semantic Web technology allows for different usage scenarios, too. We can think of specialised OntoMedia Ontologies for Health Care scenarios, where special extractors index metadata of medical information systems and libraries. Medical staff could use OntoMedia to bridge the semantic gap between different medical databases using a uniform vocabulary to search for documents relevant to specific diseases.

As the manual annotation of documents with metadata is very time consuming, the critical point of the metadata searches in general, and OntoMedia in special, is the quality and amount of metadata available in the documents. If we examine the metadata of MP3-format music, we know that plenty of files are annotated, but the annotation's quality may be poor. Semantic Web technology in itself is not capable to correct the spelling of an artist's name. But along with other technology (e.g., audio fingerprinting to identify music by content analysis) Semantic Web technology provides a standardised platform to describe, exchange, query, and reason about metadata from documents.

CONCLUSION AND OUTLOOK

In this paper, we introduced the application "OntoMedia" which uses Semantic Web technology to implement an ontology driven personal multimedia information system. The main goal of this system was the management of large multimedia collections using semantic integration techniques for metadata by applying state of the art ontology driven Semantic Web technology to the multimedia domain. As compared to competitive systems, OntoMedia provides the following core advantages:

- Semantic metadata integration using standardised ontology languages to define semantic mappings between different ontologies and proprietary formats.
- Semantic query evaluation using inference techniques based on RDFS semantics.
- An innovative graphical user interface using poly hierarchies to organise categorisation structures which provide multiple classifications based on standardised RDFS semantics.
- An overall extensible platform independent architecture which allows for easy data exchange interfaces using RDF as an extensible base for both semantic schema information and instance data.

OntoMedia was built using platform independent Java Technology, which provides an easy test ride of the application without complicated installation procedures using the Java WebStart access point at http://www.ontomedia.de/webstart.

Future versions of OntoMedia will include better access to external metadata sources providing by web databases such as Musicbrainz (http://www.muscibrainz.org), Internet Movie Database (http://www.imdl.org), or FreeDB (http://www.freedb.org). In addition, the current prototype is not well tuned to help personalisation of the presented metadata. We will include usage feedback mechanisms to adapt the presentation of the category graph to user behaviour by statistical analysis. Doing this, OntoMedia will calculate often used navigation paths in the categorisation graph and adapt its presentation of the graph. In addition, we might include content specific indexing functionality in future versions, which will be easy for textual content but is still a problem analyzing audio visual media.

REFERENCES

Appelt, D., Hobbs, J., Bear, J., Israel, D., & Tyson, M. (1993). FASTUS: A finite-state processor for information extraction from real-world text. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI)*, Chambery, France.

Becket, D. (2004). *RDF/XML syntax specification* (W3C Recommendation).

Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May). The Semantic Web. *Scientific American*.

Bimbo, A. D. (1999). *Visual information retrieval.* Morgan Kaufmann Publishers.

Broekstra, J., Kampman, A., & van Harmelen, F. (2002). Sesame: An architecture for storing and querying RDF and RDF schema. In *Proceedings of the 1st International Semantic Web Conference (ISWC 2002), LNCS 2342*, Sardinia, Italy. Springer.

Cano, P., Batlle, E., Kalker, T., & Haitsma, J. (2005, November). A review of audio fingerprinting. *Journal of VLSI Signal Process Systems, 41*(3), 271-284.

Eades, P. (1984). A heuristic for graph drawing. *Congressus Numerantium, Band 42*.

Foote, J. T. (1997). Content-based retrieval of music and audio. In C. C. J. Kuo et al. (Eds.), *Multimedia storage and archiving systems II, Proceedings of SPIE (*Vol. 3229, pp. 138-147).

Gemmell, J., Bell, G., & Lueder, R. (2006, January). MyLifeBits: A personal database for everything. *Communications of the ACM, 49*(1), 88-95.

Handschuh, S., & Staab, S. (2003). *Annotation for the Semantic Web.* IOS Press.

Hüsemann, B., & Vossen, G. (2004). Ontology-driven multimedia object management for private users: Overview and research issues. *AIS SIGSEMIS Bulletin, 1*(1).

Hüsemann, B., & Vossen, G. (2005). Ontology engineering from a database perspective. In *Proceedings of the 10th Asian Computing Science Conference (ASIAN 2005): Data Management on the Web), LNCS 3818*, Kunming, China (pp. 49-63). Springer.

Lassila, O., & Swick, R. (2004). *Resource Description Framework (RDF) Model and syntax specification* (W3C Recommendation).

Rowe, L. A., & Jain, R. (2005). ACM SIGMM retreat report on future directions in multimedia research. *ACM Transactions on Multimedia Computing, Communications, and Applications, 1*(1), 3-13.

ADDITIONAL RESOURCES

3vix: http://www.3ivx.com/
Copernic Desktop Search: http://www.copernic.com/en/products/desktop-search
FreeDB: http://www.freedb.org
Google Desktop: http://desktop.google.com/
Java Media Framework: http://java.sun.com/products/java media/jmf/index.jsp
JLayer: http://www.javazoom.net/javalayer/javalayer.html
MSN Toolbar: http://toolbar.msn.com/
Musicbrainz: http://www.musicbrainz.org
Quicktime Java SDK: http://www.apple.com/quicktime/qtjava
Touchgraph Framework: http://www.touchgraph.com/
x-friend: http://www.x-friend.de

This work was previously published in the International Journal on Semantic Web & Information Systems, Vol. 2, Issue 3, edited by A. Sheth and M. Lytras, pp. 1-16, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.7
Semantic Multimedia Information Analysis for Retrieval Applications

João Magalhães
Imperial College London, UK

Stefan Rüger
Imperial College London, UK

ABSTRACT

Most of the research in multimedia retrieval applications has focused on retrieval by content or retrieval by example. Since the classical review by Smeulders, Worring, Santini, Gupta, and Jain (2000), a new interest has grown immensely in the multimedia information retrieval community: retrieval by semantics. This exciting new research area arises as a combination of multimedia understanding, information extraction, information retrieval, and digital libraries. This chapter presents a comprehensive review of analysis algorithms in order to extract semantic information from multimedia content. We discuss statistical approaches to analyze images and video content and conclude with a discussion regarding the described methods.

INTRODUCTION: MULTIMEDIA ANALYSIS

The growing interest in managing multimedia collections effectively and efficiently has created new research interest that arises as a combination of multimedia understanding, information extraction, information retrieval, and digital libraries. This growing interest has resulted in the creation of a video retrieval track in TREC conference series in parallel with the text retrieval track (TRECVID, 2004).

Figure 1 illustrates a simplified multimedia information retrieval application composed by a multimedia database, analysis algorithms, a description database, and a user interface application. Analysis algorithms extract features from multimedia content and store them as descriptions of that content. A user then deploys these indexing descriptions in order to search the multimedia

Figure 1. A typical multimedia information retrieval application

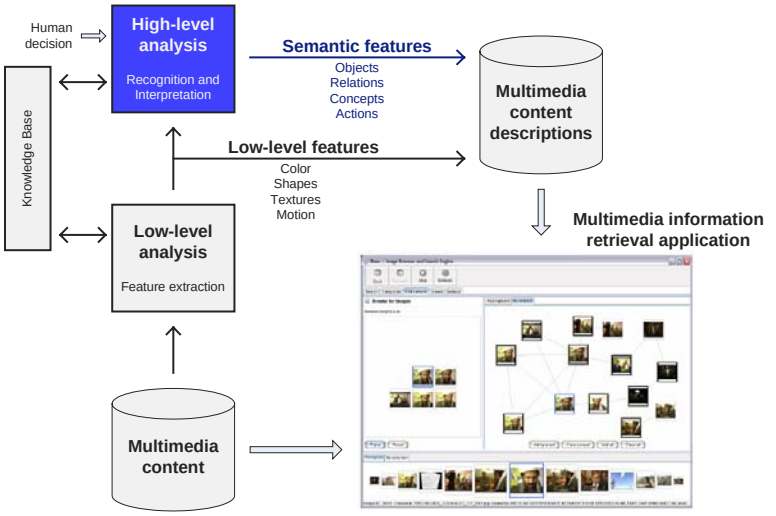
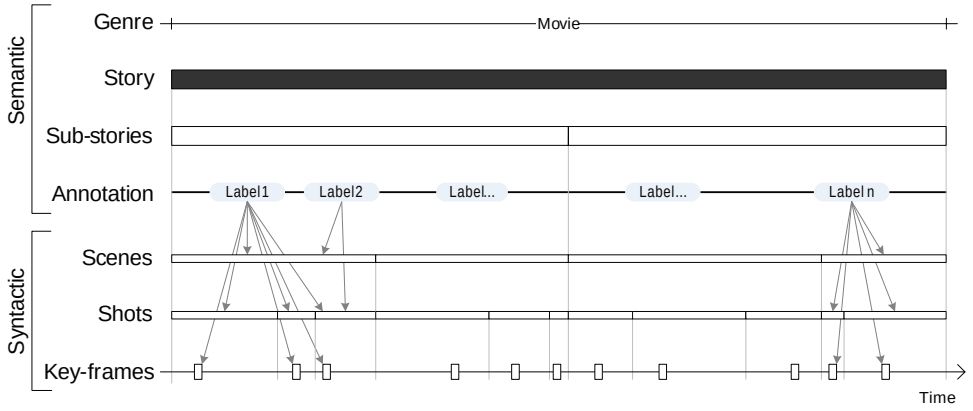


Figure 2. Syntactic and semantic structure of video



database. A semantic multimedia information retrieval application (Figure 1) differs eminently from traditional retrieval applications on the low-level analysis algorithms; its algorithms are responsible for extracting semantic information used to index multimedia content by its semantic. Multimedia content can be indexed in many ways, and each index can refer to different modalities and/or parts of the multimedia piece. Multimedia content is composed of the visual track, sound track, speech track, and text. All these modalities are arranged temporally to provide a meaningful way to transmit information and/or entertainment.

The way video documents are temporally structured can be distinguished in two levels: semantic and syntactic structure (Figure 2).

At the syntactic level, the video is segmented into shots (visual or audio) that form a uniform segment (e.g., visually similar frames); representative key-frames are extracted from each shot, and scenes group neighboring similar shots into a single segment. The segmentation of video into its syntactic structure of video has been studied widely (Brunelli, Mich, & Modena, 1999; Wang, Liu, & Huang, 2000).

At the semantic level, annotations of the key-frames and shots with a set of labels indicate the presence of semantic entities, their relations, and attributes (agent, object, event, concept, state, place, and time (see Benitez et al., 2002, for details). Further analysis allows the discovery of logical sub-units (e.g., substory or subnarrative), logical units (e.g., a movie), and genres. A recent review of multimedia semantic indexing has been published by Snoek and Worring (2005).

The scope of this chapter is the family of semantic-multimedia analysis algorithms that automate the multimedia semantic annotation process. In the following sections, we will review papers on multimedia-semantic analysis: semantic annotation of key-frame images, shots, and scenes. The semantic analysis at the shot and scene level considers independently the audio and visual modalities and then the multi-modal semantic analysis. Due to the scope of this book, we will give more emphasis to the visual part than to the audio part of the multimedia analysis and will not cover the temporal analysis of logical substories, stories, and genres.

KEY-FRAME SEMANTIC ANNOTATION

Image analysis and understanding is one of the oldest fields in pattern recognition and artificial intelligence. A lot of research has been done since (Marr, 1983), culminating in the modern reference texts by Forsyth and Ponce (2003) and Hartley and Zisserman (2004). In the following sections we discuss different types of visual information analysis algorithms: single class models fit a simple probability density distribution to each label; translation models define a visual vocabulary and a method to translate from this vocabulary to keywords; hierarchical and network models explore the interdependence of image elements (regions or tiles) and model its structure; knowledge-based models improve the model's accuracy by including other sources of knowledge besides the training data (e.g., a linguistic database such as WordNet).

Single Class Models

A direct approach to the semantic analysis of multimedia is to learn a class-conditional probability distribution $p(w \mid x)$ of each single keyword w of the semantic vocabulary, given its training data x (see Figure 3). This distribution can be obtained by using Bayes' law:

$$p(w \mid x) = \frac{p(x \mid w)\, p(w)}{p(x)}.$$

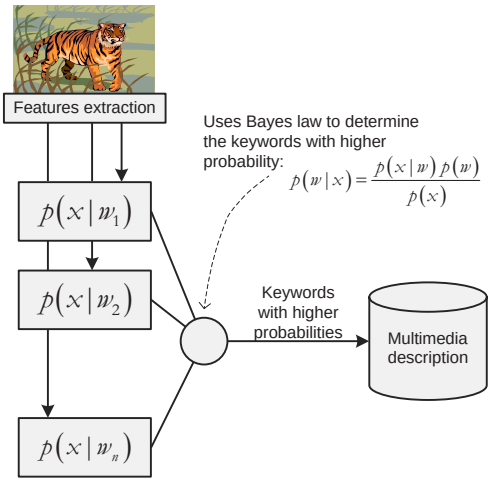
The data probability $p(x)$ and the keyword probability $p(w)$ can be computed straightforward, and the $p(x \mid w)$ can be computed with very different data density distribution models.

Several techniques to model the $p(x \mid w)$ with a simple density distribution have been proposed: Yavlinsky, Schofield, and Rüger (2005) used a nonparametric distribution, Carneiro and Vasconcelos (2005) a semi-parametric density estimation, Westerveld and de Vries (2003) a finite mixture of Gaussians, and Mori, Takahashi, and Oka (1999), Vailaya, Figueiredo, Jain, and Zhang (1999), and Vailaya, Figueiredo, Jain, and Zhang (2001) different flavors of vector quantization techniques.

Yavlinsky et al. (2005) modeled the probability density of images, given keywords as a nonparametric density smoothed by two kernels: a Gaussian kernel and an Earth Mover's Distance kernel. They used both global and 3×3 tile color and texture features. The best reported mean average precision (MAP) results with tiles achieved 28.6% MAP with the dataset of Duygulu, Barnard, de Freitas, and Forsyth (2002) and 9.2% with a Getty Images dataset.

Yavlinsky et al. (2005) showed that a simple nonparametric statistical distribution can perform as well or better than many more sophisticated techniques (e.g., translation models). However, the

Figure 3. Inference of single class models



nonparametric density nature of their framework makes the task of running the model on new data very complex. The model is the entire dataset meaning that the demands on CPU and memory increase with the training data.

Westerveld and de Vries (2003) used a finite-mixture density distribution with a fixed number of components to model a subset of the DCT coefficients:

$$p(x|\theta)=\sum_{m=1}^{k}\alpha_m p(x|\mu_m,\sigma_m^2),$$

in which k is the number of components, θ represents the complete set of model parameters with mean μ_m, covariance σ_m^2, and component prior α_m. The component priors have the constraints $\alpha_1,...,\alpha_k \geq 0$ and $\sum_{m=1}^{k}\alpha_m = 1$. Westerveld (2003) tested several scenarios to evaluate the effect (a) of the number of mixture components, (b) of using different numbers of DCT coefficients (luminance and chrominance), and (c) of adding the coordinates of the DCT coefficients to the feature vectors. The two first factors produced varying results, and optimal points were found experimentally. The third tested aspect, the presence of the coefficients position information, did not modify the results.

Marrying the two previous approaches, Carneiro and Vasconcelos (2005) deployed a hierarchy of semi-parametric mixtures to model $p(x \mid w)$ using a subset of the DCT coefficients as low-level features. Vasconcelos and Lippman (2000) had already examined the same framework in a content-based retrieval system.

The hierarchy of mixtures proposed by Vasconcelos and Lippman (1998) can model data at different levels of granularity with a finite mixture of Gaussians. At each hierarchical level *l*, the number of each mixture component k^l differs by one from adjacent levels. The hierarchy of mixtures is expressed as:

$$p(x|w_i)=\frac{1}{D}\sum_{m=1}^{k^l}\alpha_{i,m}^{l}p(x|\theta_{i,m}^{l}).$$

The level *l=1* corresponds to the coarsest characterization. The more detailed hierarchy level consists of a nonparametric distribution with a kernel placed on top of each sample. The only restriction on the model is that if node *m* of level *l+1* is a child of node *n* of level *l*, then they are both children of node *p* of level *l-1*. The EM algorithm computes the mixture parameters at level *l*, given the knowledge of the parameters at level *l+1*, forcing the previous restriction.

Carneiro and Vasconcelos (2005) report the best published retrieval MAP of 31% with the dataset of Duygulu et al. (2002). Even though we cannot dissociate this result from the pair of features and statistical model, the hierarchy of mixtures appears to be a very powerful density distribution technique.

Even though the approaches by Carneiro and Vasconcelos (2005) and Westerveld and de Vries (2003) are similar, the differences make it difficult to do a fair comparison. The DCT features are used in a different way, and the semi-parametric hierarchy of mixtures can model classes with very few training examples.

The relationship between finite-mixture density modeling and vector quantization is a well-studied subject (see Hastie, Tibshirani, &

Friedman, 2001). One of the applications of vector quantization to image retrieval and annotation was realized by Mori et al. (1999). Given the training data of a keyword, they divide the images into tiles and apply vector quantization to the image tiles in order to extract the codebook used to estimate the $p(x \mid w)$ density distribution. Later, they use a model of word co-occurrence on the image tiles in order to label the image. The words with the higher sum of probabilities across the different tiles are the ones assigned to that image.

Vailaya et al. (1999) and Vailaya et al. (2001) describe a Bayesian framework using a codebook to estimate the density distribution of each keyword. They show that the Minimum Description Length criterion selects the optimal size of the codebook extracted from the vector quantizer. The features are extracted from the global image, and there is no image tiling. The use of the MDL criterion makes this framework quite elegant and defines a statistical criterion to select every model parameter and without any user-defined parameters.

Translation Models

All of the previous approaches employ a direct model to estimate $p(x \mid w)$ with image global features and/or image tiles features. In contrast to this, the vector quantization (usually k-means) approach generates a codebook of image regions or image tiles (depending on the segmentation solution). The problem then is formulated as a translation problem between two representations of the same entity: English-Esperanto, word-blob codebook, or word-tile codebook.

Inspired by machine translation research, Duygulu et al. (2002) developed a method of annotating image regions with words. First, regions are created using a segmentation algorithm like normalized cuts (Shi & Malik, 2000). For each region, features are computed, and then blobs are generated by clustering the regional image features across an image collection. The problem then is formulated as learning the correspondence between the discrete vocabulary of blobs and the image keywords. The model consists of a mixture of correspondences for each word of each image in the collection:

$$p\left(w_j \mid I_n\right) = \sum_{i \in \{\text{blobs in } I_n\}} p\left(a_{nj} = i\right) p\left(w = w_{nj} \mid b = b_i\right),$$

in which $p(a_{nj} = i)$ expresses the probability of associating word j to blob i in image n, and $p(w = w_{nj} \mid b = b_i)$ is the probability of obtaining an instance of word w given an instance of blob b. These two probability distributions are estimated with the EM algorithm. The authors refined the lexicon by clustering indistinguishable words and ignoring the words with probabilities $p(w \mid b)$ below a given threshold.

The machine translation approach, the thorough experiments, and the dataset form strong points of this chapter (Duygulu et al., 2002). This dataset is nowadays a reference, and thorough experiments showed that (a) their method could predict numerous words with high accuracy, (b) increasing the probability threshold improved precision but reduced recall, and (c) the word clustering improved recall and precision.

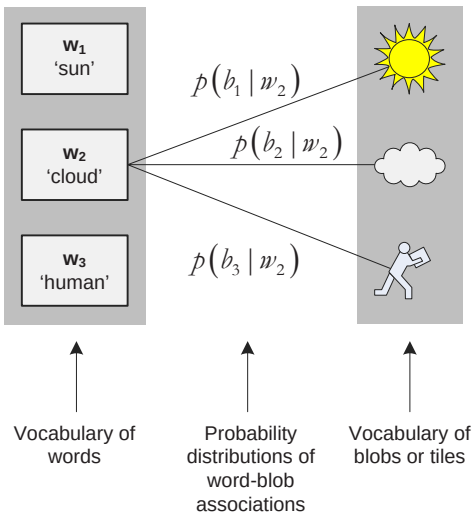
Following a translation model, Jeon, Lavrenko, and Manmatha (2003), Lavrenko, Manmatha, and Jeon (2003), and Feng, Lavrenko, and Manmatha (2004) studied a model in which blob features $b_I^{(r)}$ of an image I are assumed to be conditionally independent of keywords w_i, that is:

$$\begin{aligned} p\left(w_i, b_I\right) &= \sum_{J \in D} P(J) P\left(w_i \mid J\right) P\left(b_I \mid J\right) \\ &= \sum_{J \in D} P(J) P\left(w_i \mid J\right) \prod_{r \in I} P\left(b_I^{(r)} \mid J\right) \end{aligned}$$

Note that $b_I^{(r)}$ and w_i are conditionally independent, given the image collection D and that $J \in D$ act as the hidden variables that generated the two distinct representations of the same process (words and features).

Jeon et al. (2003) recast the image annotation as a cross-lingual information retrieval problem,

Figure 4. Translation models



applying a cross-media relevance model based on a discrete codebook of regions. Lavrenko et al. (2003) continued their previous work (Jeon et al., 2003) and used continuous probability density functions $p(b_l^{(r)} \mid J)$ to describe the process of generating blob features and to avoid the loss of information related to the generation of the codebook. Extending their previous work, Feng et al. (2004) replaced blobs with tiles and modeled image keywords with a Bernoulli distribution. This last work reports their best results, a MAP of 30%, with a Corel dataset (Duygulu et al., 2002).

Latent semantic analysis is another technique of text analysis and indexing; it looks at patterns of word distributions (specifically, word co-occurrence) across a set of documents (Deerwester, Dumais, Furmas, Landauer, & Harshman, 1990). A matrix *M* of word occurrences in documents is filled with each word frequency in each document. The singular value decomposition (SVD) of matrix *M* gives the transformation to a singular space in which projected documents can be compared efficiently.

Hierarchical Models

The aforementioned approaches assumed a minimal relation among the various elements of an image (blobs or tiles). This section and the following section will review methods that consider a hierarchical relation or an interdependence relation among the elements of an image (words and blobs or tiles).

Barnard and Forsyth (2001) studied a generative hierarchical aspect model, which was inspired by Hofmann and Puzicha's (1998) hierarchical clustering/aspect model. The data are assumed to be generated by a fixed hierarchy of nodes in which the leaves of the hierarchy correspond to soft clusters. Mathematically, the process for generating the set of observations O associated with an image I can be described by Box 1, in which c indexes the clusters, o indexes words and blobs, and l indexes the levels of the hierarchy. The level and the cluster uniquely specify a node of the hierarchy. Hence, the probability of an observation $p(o \mid l, c)$ is conditionally independent given a node in the tree. In the case of words, $p(o \mid l, c)$ assumes a tabular form, and in the case of blobs, a single Gaussian models the regions' features. The model is estimated with the EM algorithm.

Blei and Jordan (2003) describe three hierarchical mixture models to annotate image data, culminating in the correspondence latent Dirichlet allocation model. It specifies the following joint distribution of regions, words, and latent variables (θ, z, y):

Box 1.

$$p(O \mid I) = \sum_c \left(p(c) \prod_{o \in O} \left(\sum_l p(o \mid l, c)\, p(l \mid c, I) \right) \right), \quad O = \{w_1, \ldots, w_n, b_1, \ldots, b_m\}$$

$$p(r,w,\theta,z,y)=p(\theta\mid\alpha)\left(\prod_{n=1}^{N}p(z_n\mid\theta)p(r_n\mid z_n,\mu,\sigma)\right)\cdot\left(\prod_{m=1}^{M}p(y_m\mid N)p(w_m\mid y_m,z,\beta)\right).$$

This model assumes that a Dirichlet distribution θ (with α as its parameter) generates a mixture of latent factors: z and y. Image regions r_n are modeled with Gaussians with mean μ and covariance σ, in which words w_n follow a multinomial distribution with a β parameter.

This mixture of latent factors then is used to generate words (y variable) and regions (z variable). The EM algorithm estimates this model, and the inference of $p(w\mid r)$ is carried out by variational inference. The correspondence latent Dirichlet allocation model provides a clean probabilistic model for annotating images with multiple keywords. It combines the advantages of probabilistic clustering for dimensionality reduction with an explicit model of the conditional distribution from which image keywords are generated.

Li and Wang (2003) characterize the images with a hierarchical approach at multiple tiling granularities (i.e., each tile in each hierarchical level is subdivided into smaller sub-tiles). A color and texture feature vector represents each tile. The texture features represent the energy in high-frequency bands of wavelet transforms. They represent each keyword separately with two-dimensional, multi-resolution hidden Markov models. This method achieves a certain degree of scale invariance due to the hierarchical tiling process and the two-dimensional multiresolution hidden Markov model.

Network Models

In semantic-multimedia analysis, concepts are interdependent; for example, if a house is detected in a scene, then the probability of existing windows and doors in the scene are boosted, and vice-versa. In other words, when inferring the probability of a set of interdependent random variables, their probabilities are modified iteratively until an optimal point is reached (to avoid instability, the loops must exist over a large set of random variables [Pearl, 1988]). Most of the papers discussed next model keywords as a set of interdependent random variables connected in a probabilistic network.

Various graphical models have been implemented in computer vision to model the appearance, spatial relations, and co-occurrence of local parts. Markov random fields and hidden Markov models are the most common generative models that learn the joint probability of the observed data (X) and the corresponding labels (Y). These models divide the image into tiles or regions (other approaches use contour directions, but these are outside the scope of our discussion). A probabilistic network then models this low-level division in which each node corresponds to one of these tiles or regions and its label. The relation among nodes depends on the selected neighboring method. Markov random fields can be expressed as:

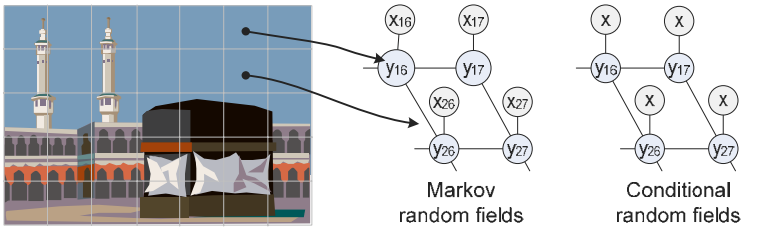
$$P(x,y)=\frac{1}{Z}\cdot\prod_{i}\left(\phi_i(x_i,y_i)\cdot\prod_{j\in N_i}\varphi_{i,j}(y_i,y_j)\right),$$

in which i indexes the image's tiles, j indexes the neighbors of the current i tile, ϕ_i is the potential function of the current tile x_i, and its possible labels y_i, and $\varphi_{i,j}$ are the interaction functions between the current tile label and its neighbors. Figure 5 illustrates the Markov random field framework.

The Markov condition implies that a given node only depends on its neighboring nodes. This condition constitutes a drawback for these models, because only local relationships are incorporated into the model. This makes it highly unsuitable for capturing long-range relations or global characteristics.

In order to circumvent this limitation, Kumar and Herbert (2003a) propose a multi-scale random field (MSRF) as a prior model on the class

Figure 5. Two types of random fields



labels on the image sites. This model implements a probabilistic network that can be approximated by a 2D hierarchical structure such as a 2D-tree. A multiscale feature vector captures the local dependencies in the data. The distribution of the multiscale feature vectors is modeled as a mixture of Gaussians. The features were selected specifically to detect human-made structures, which are the only types of objects that are detected.

Kumar and Herbert's (2003) second approach to this problem is based on discriminative random fields, an approach inspired on conditional random fields (CRF). CRFs, defined by Lafferty, McCallum, and Pereira (2001), are graphical models, initially for text information extraction, that are meant for visual information analysis in this approach. More generally, a CRF is a sequence-modeling framework based on the conditional probability of the entire sequence of labels (Y), given the all image (X). CRFs have the following mathematical form:

$$P(y|x)=\frac{1}{Z}\cdot\prod_{i}\left(\phi_i(y_i,x)\cdot\prod_{j\in N_i}\varphi_{i,j}(y_i,y_j;x)\right),$$

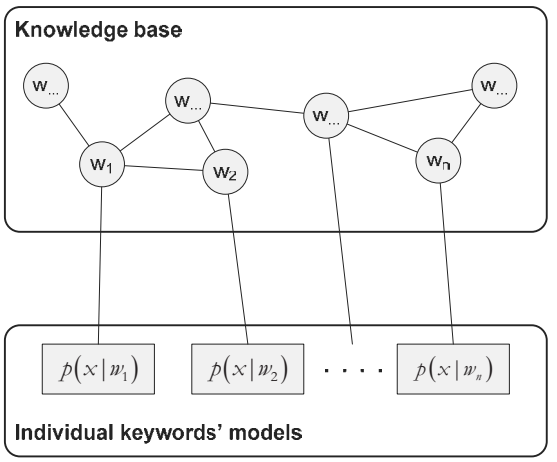
in which i indexes the image's tiles, j indexes the neighbors of the current i tile, ϕ_i is the association potential between the current tile and the image label, and $\varphi_{i,j}$ is the interaction potential between the current tile and its neighbors (note that it is also dependent on the image label). Figure 5 illustrates the conditional random field framework. The authors showed that this last approach outperformed their initial proposal of a multiscale random field as well as the more traditional MRF solution in the task of detecting human-made structures.

He, Zemel, and Carreira-Perpiñán (2004) combine the use of a conditional random field and data at multiple scales. Their multiscale conditional random field (mCRF) is a product of individual models, each model providing labeling information from various aspects of the image: a classifier that looks at local image statistics; regional label features that look at local label patterns; and global label features that look at large, coarse label patterns. The mCRF is shown to detect several types of concepts (i.e., sky, water, snow, vegetation, ground, hippopotamus, and bear) with classification rates better than a traditional Markov random field.

Quattoni, Collins, and Darrell (2004) extend the CRF framework to incorporate hidden variables and combine class-conditional CRFs into a unified framework for part-based object recognition. The features are extracted from special regions that are obtained with the scale-invariant feature transform or SIFT (Lowe, 1999). The SIFT detector finds points in locations at scales in which there is a significant amount of variation. Once a point of interest is found, the region around it is extracted at the appropriate scale. The features from this region then are computed and plugged into the CRF framework. The advantage of this method is that it needs fewer regions by eliminating redundant regions and selecting the ones with more energy on high-frequency bands.

One should note that all these approaches require a ground truth at the level of the image's

Figure 6. Knowledge-based models



tiles/regions as is common in computer vision. This is not what is found traditionally in multimedia information retrieval datasets in which the ground truth exists rather at a global level.

Knowledge-Based Models

The previous methods have only visual features as training data to create the statistical models in the form of a probabilistic network. Most of the time, these training data are limited, and the model's accuracy can be improved by other sources of knowledge. Prior knowledge can be added to a model either by a human expert who states the relations between concept variables (nodes in a probabilistic network) or by an external knowledge base in order to infer the concept relations (e.g., with a linguistic database such as WordNet) (Figure 6).

Tansley (2000) introduces a multimedia thesaurus in which media content is associated with appropriate concepts in a semantic layer composed by a network of concepts and their relations. The process of building the semantic layer uses Latent Semantic Indexing to connect images to their corresponding concepts, and a measure of each correspondence (image concept) is taken from this process. After that, unlabeled images (test images) are annotated by comparing them with the training images using a *k*-nearest-neighbor classifier. Since the concepts' interdependences are represented in the semantic layer, the concepts' probability computed by the classifier are modified by the others concepts.

Other authors have explored not only the statistical interdependence of context and objects but also have used other knowledge that is not present in multimedia data, which humans use to understand (or predict) new data. Srikanth, Varner, Bowden, and Moldovan (2005) incorporated linguistic knowledge from WordNet (Miller, 1995) in order to deduce a hierarchy of terms from the annotations. They generate a visual vocabulary based on the semantics of the annotation words and their hierarchical organization in the WordNet ontology.

Benitez and Chang (2002) and Benitez (2005) took this idea further and suggested a media ontology (MediaNet) to help to discover, summarize, and measure knowledge from annotated images in the form of image clusters, word senses, and relationships among them. MediaNet, a Bayesian network-based multimedia knowledge representation framework, is composed by a network of concepts, their relations, and media exemplifying concepts and relationships. The MediaNet integrates classifiers in order to discover statistical relationships among concepts. WordNet is used to process image annotations by stripping out unnecessary information. The summarization process implements a series of strategies to improve the images' description qualities, for example using WordNet and image clusters to disambiguate annotation terms (images in the same clusters tend to have similar textual descriptions). Benitez (2005) also proposes a set of measures to evaluate the knowledge consistency, completeness, and conciseness.

Tansley (2000) used a network at the concept level, and Benitez (2005) used the MediaNet network to capture the relations at both concept and feature levels. In addition, Benitez (2005) utilized

WordNet, which captures human knowledge that is not entirely present in multimedia data.

Summary

The described algorithms vary in many different aspects such as in their low-level features, segmentation methods, feature representation, modeling complexity, or required data. While some concepts require a lot of data to estimate its model (e.g., a car), others are very simple and require just a few examples (e.g., sky). So, we advocate that different approaches should be used for different concept complexities.

Single-class models assume that concepts are independent and that each concept has its own model. These are the simplest models that can be used and the ones with better accuracy (e.g., Yavlinsky et al., 2005).

Translation models, hierarchical models, and network models capture a certain degree of the concept's interdependence (co-occurrence) from the information present in the training data. The difference between the models is linked to the degree of interdependence that can be represented by the model. In practice, when interdependencies information is incorporated in the model, it also inserts noise in the form of false interdependencies, which causes a decrease in performance. So, the theoretical advantage of these models is in practice reduced by this effect.

All these models rely exclusively on visual low-level features in order to capture complex human concepts and to correctly predict new unlabeled data. Most of the time, the training data are limited, and the model's accuracy can be improved by other sources of knowledge. Srikanth et al. (2005) and Benitez (2005) are two of the few proposals that exploit prior knowledge that is external to the training data in order to capture the interdependent (co-occurrence) nature of concepts.

At this time, knowledge-based models seem to be the most promising semantic analysis algorithms for information retrieval. Text information retrieval already has shown great improvement over exclusively statistical models when external linguistic knowledge was used (Harabagiu et al., 2000). Multimedia retrieval will go through a similar progress but at a slower pace, because there is no multimedia ontology that offers the same knowledge base as WordNet offers to linguistic text processing.

SHOT AND SCENE SEMANTIC ANNOTATION

Shot and scene semantic analysis introduces the time dimension to the problem at hand. The time dimension adds temporal frames, resulting in more information to help the analysis. To take advantage of the sequential nature of the data, the natural choices of algorithms are based on hierarchical models or network models. The section is organized by modality, and within each modality, we don't detail the algorithms by technique due to space constraints. This way, we shed some light on multimodality shot and scene semantic analysis and keep the chapter's emphasis on visual information analysis.

Audio Analysis

Audio analysis becomes a very important part of the multimodal analysis task when processing TV news, movies, sport videos, and so forth. Various types of audio can populate the sound track of a multimedia document, the most common types being speech, music, and silence. Lu, Zhang, and Jiang (2002) propose methods to segment audio and to classify each segment as speech, music, silence, and environment sound. A *k*-nearest neighbor model is used at the frame level followed by vector quantization to discriminate between speech and nonspeech. A set of threshold-based rules is used in order to discriminate among silence, music, and environment sound. The authors also describe a speaker change detection algorithm

based on Gaussian-mixture models (GMM); this algorithm continuously compares the model of the present speaker's speech with a model that is created dynamically from the current audio frame. After a speaker change has been detected, the new GMM replaces the current speaker's GMM.

In most TV programs and sport videos, sound events do not overlap, but in narratives (movies and soap operas), these events frequently occur simultaneously. To address this problem, Akutsu, Hamada, and Tonomura (1998) present an audio-based approach to video indexing by detecting speech and music independently, even when they occur simultaneously. Their framework is based on a set of heuristics over features histograms and corresponding thresholds. With a similar goal, Naphade and Huang (2000) define a generic statistical framework based on hidden Markov models (Rabiner, 1989) in order to classify audio segments into speech, silence, music, and miscellaneous and their co-occurrences. By creating an HMM for each class and every combination of classes, the authors achieved a generic framework that is capable of modeling various audio events with high accuracy.

Another important audio analysis task is the classification of the musical genre of a particular audio segment. This can capture the type of emotion that the director wants to communicate (e.g., stress, anxiety, happiness). Tzanetakis and Cook (2002) describe their work on categorizing music as rock, dance, pop, metal, classical, blues, country, hip-hop, reggae, or jazz (jazz and classical music had more subcategories). In addition to the traditional audio features, they also use special features to capture rhythmic characteristics and apply simple statistical models such as GMM and KNN to model each class' feature histogram. Interestingly, the best reported classification precision (61%) is in the same range as human performance for genre classification (70%).

All these approaches work as a single class model of individual classes/keywords. Note that the hidden Markov model is, in fact, a probabilistic network for modeling a single temporal event that corresponds to a given concept/keyword. So, even though it is a network model, it is used as a single class model.

Visual Analysis

Many of the visual video analysis methods are based on heuristics that are deduced empirically. Statistical methods are more common when considering multimodal analysis. Most of the following papers explore the temporal evolution of features to semantically analyze video content (e.g., shot classification, logical units, etc.). Video visual analysis algorithms are of two types: (a) heuristics-based, in which a set of threshold rules decides the content class, and (b) statistical algorithms that are similar to the ones described in Section 2.

Heuristic methods rely on deterministic rules that were defined in some empirical way. These methods monitor histograms, and events are detected if the histogram triggers a given rule (usually a threshold). They are particularly adequate for sport videos because broadcast TV follows a set of video production rules that result in well-defined semantic structures that ease the analysis of the sports videos. Several papers have been published on sports video analysis, such as football, basketball and tennis, in order to detect semantic events and to semantically classify each shot (Li & Sezan, 2003; Luo & Huang, 2003; Tan, Saur, Kulkarni, & Ramadge, 2000).

Tan et al. (2000) introduced a model for estimating camera movements (pan, tilt, and zoom) from the motion vectors of compressed video. The authors further showed how camera motion histograms could be used to discriminate various basketball shots. Prior to this, the video is segmented into shots based on the evolution of the intensity histogram across different frames. Shots are detected if the histogram exceeds a predefined threshold; then, they are discriminated based on (a) the accumulated histogram of cam-

era motion direction (fast breaks and full-court advances), (b) the slope of this histogram (fast breaks or full-court advances), (c) sequence of camera movements (shots at the basket), and (d) persistence of camera motion (close-ups).

Other heuristic methods deploy color histograms, shot duration, and shot sequences to automatically analyze various types of sports such as football (Ekin, Tekalp, & Mehrotra, 2003) and American football (Li & Sezan, 2003).

The statistical approaches reviewed previously can be applied to the visual analysis of video content with the advantage that shapes obtained by segmentation are more accurate due to the time dimension. Also, analyzing several key-frames of the same shot and then combining the results facilitate the identification of semantic entities in a given shot.

Luo and Hwang's (2003) statistical framework tracks objects within a given shot with a dynamic Bayesian network and classifies that shot from a coarse-grain to a fine-grain level. At the course-grain level, a key-frame is extracted from a shot every 0.5 seconds. From these key-frames, motion and global features are extracted, and their temporal evolution is modeled with a hierarchical hidden Markov model (HHMM). Individual HHMMs (a single-class model approach) capture a given semantic shot category. At the fine-grain level analysis, Luo and Hwang (2003) employ object recognition and tracking techniques. After the coarse-grain level analysis, segmentation is performed on the shots to extract visual objects. Then, invariant points are detected in each shape to track the object movement. These points are fed to a dynamic Bayesian network to model detailed events occurring within the shot (e.g., human body movements in a golf game).

Souvannavong, Merialdo, and Huet (2003) used latent semantic analysis to analyze video content. Recall that latent semantic analysis algorithm builds a matrix M of word occurrences in documents, and then the SVD of this matrix is computed to obtain a singular space. The problem with multimedia content is that there is no text corpus (a vocabulary). A vector quantization technique (k-means) returns a codebook of blobs, the vocabulary of blobs from the shots' key-frames. In the singular feature space, a k-nearest-neighbor ($k=20$) and a Gaussian mixture model technique are used to classify new videos. The comparison of the two techniques shows that GMM performs better when there is enough data to correctly estimate the 10 components. The k-nn algorithm has the disadvantages of every nonparametric method—the model is the training data, and for the TRECVID dataset (75,000 key-frames), training can take considerable time.

Multimodal Analysis

In the previous analysis, the audio and visual modalities were considered independently in order to detect semantic entities. These semantic entities are represented in various modalities, capturing different aspects of that same reality. Those modalities contain co-occurring patterns that are synchronized in a given way because they represent the same reality. Thus, synchronization and the strategy to combine the multimodal patterns is the key issue in multimodal analysis. The approaches described in this section explore the multimodality statistics of semantic entities (e.g., pattern synchronization).

Sports video analysis can be greatly improved with multimodal features; for example, the level of excitement expressed by the crowd noise can be a strong indicator of certain events (foul, goal, goal miss, etc). Leonardi, Migliotari, and Prandini (2004) take this into account when designing a multimodal algorithm to detect goals in football videos. A set of visual features from each shot is fed to a Markov chain in order to evaluate their temporal evolution from one shot to the next. The Markov chain has two states that correspond to the goal state and to the nongoal state. The visual analysis returns the positive pair shots, and the shot audio loudness is the criterion to rank the

pair shots. Thus, the two modalities never are combined but are used sequentially. Results show that audio and visual modalities together improve the average precision when compared only to the audio case (Leonardi et al., 2004).

In TV news videos, text is the fundamental modality with the most important information. Westerveld, et al. (2003) build on their previous work described previously to analyze the visual part and to add text provided by an Automatic Speech Recognition (ASR) system. The authors further propose a visual dynamic model to capture the visual temporal characteristics. This model is based on the Gaussian mixture model estimated from the DCT blocks of the frames around each key-frame in the range of 0.5 seconds. In this way, the most significant moving regions are represented by this model with an evident applicability to object tracking. The text retrieval model evaluates a given $Shot_i$ for the queried keywords $Q = q_1, q_2, q_3, ...\}$ (see Box 2).

This measure evaluates the probability that one or more queried keywords appear in the evaluated shot, $p(q_k \mid Shot_i)$, or in the scene, $p(q_k \mid Scene_i)$, under the prior $p(q_k)$. The λ variables correspond to the probabilities of corresponding weights. This function, inspired by language models, creates the scene-shot structure of video content. The visual model and the text model are combined under the assumption that they are independent; thus, the probabilities are simply multiplied. The results with both modalities are reported to be better than using just one.

Naphade and Huang (2001) characterize single-modal concepts (e.g., indoor/outdoor, forest, sky, water) and multimodal concepts (e.g., explosions, rocket launches) with Bayesian networks. The visual part is segmented into shots (Naphade et al., 1998), and from each key-frame, a set of low-level features is extracted (color, texture, blobs, and motion). These features then are used to estimate a Gaussian mixture model of multimedia concepts at region level and then at frame level. The audio part is analyzed with the authors' algorithm described previously (Naphade & Huang, 2000). The outputs of these classifiers are then combined in a Bayesian network in order to improve concept detection. Their experiments show that the Bayesian network improves the detection performance over individual classifiers. IBM's research by Adams et al. (2003) extend the work of Naphade and Huang (2001) by including text from Automatic Speech Recognition as a third modality and by using Support Vector Machines to combine the classifiers' outputs. The comparison of these two combination strategies showed that SVMs (audio, visual, and text) and Bayesian networks (audio and visual) perform equally well. However, since in the latter case, speech information was ignored, one might expect that Bayesian networks can, in fact, perform better. More details about IBM's research work can be found in Naphade and Smith (2003), Natsev, Naphade, and Smith (2003), and Tseng, Lin, Naphade, Natsev, and Smith (2003).

The approach by Snoek and Worring (2005) is unique in the way synchronization and time relations between various patterns are modeled explicitly. They propose a multimedia semantic analysis framework based on Allen's (1983) temporal interval relations. Allen showed that in order to maintain temporal knowledge about any two events, only a small set of relations is needed to represent their temporal relations. These relations, now applied to audio and visual patterns, are the following: precedes, meets, overlaps, starts, dur-

Box 2.

$$RSV\left(Shot_i\right)=\frac{1}{|Q|}\sum_{k=1}^{|Q|}\log\left(\lambda_{Shot}\,p\left(q_k \mid Shot_i\right)+\ _{Scene}\,p\left(q_k \mid Scene_i\right)+\lambda_{Coll}\,p\left(q_k\right)\right)$$

ing, finishes, equals, and no relation. The framework can include context and synchronization of heterogeneous information sources involved in multimodal analysis. Initially, the optimal pattern configuration of temporal relations of a given event is learned from training data by a standard statistical method (maximum entropy, decision trees, and SVMs). New data are classified with the learned model. The authors evaluate the event detection on a soccer video (goal, penalty, yellow card, red card and substitution) and TV news (reporting anchor, monologue, split-view and weather report). The differences among the various classifiers (maximum entropy, decision trees, and SVMs) appear to be not statistically significant.

Summary

When considering video content, a new, very important dimension is added: time. Time adds a lot of redundancy that can be explored effectively in order to achieve a better segmentation and semantic analysis. The most interesting approaches consider time either implicitly (Westerveld et al., 2003) or explicitly (Snoek & Worring, 2005).

Few papers show a deeper level of multimodal combination than Snoek and Worring (2005) and Naphade and Huang (2001). The first explicitly explores the multimodal co-occurrence of patterns resulting from the same event with temporal relations. The latter integrates multimodal patterns in a Bayesian network to explore pattern co-occurrences and concept interdependence.

Natural language processing experts have not yet applied all the techniques from text to the video's extracted speech. Most approaches to extract information from text and combine this with the information extracted from audio and video are all very simple, such as a simple product between the probabilities of various modalities' classifiers.

CONCLUSION

This chapter reviewed semantic-multimedia analysis algorithms with special emphasis on visual content. Multimedia datasets are important research tools that provide a means for researchers to evaluate various information extraction strategies. The two parts are not separate, because algorithm performances are intrinsically related to the dataset on which they are evaluated.

Major developments in semantic-multimedia analysis algorithms will probably be related to knowledge-based models and multimodal fusion algorithms. Future applications might boost knowledge-based model research by enforcing a limited application domain (i.e., a constrained knowledge base). Examples of such applications are football game summaries and mobile photo albums.

Multimodal analysis algorithms already have proven to be crucial in semantic multimedia analysis. Large developments are expected in this young research area due to the several problems that wait to be fully explored and to the TRECVID conference series that is pushing forward this research area through a standard evaluation and a rich multimedia dataset.

We believe that semantic-multimedia information analysis for retrieval applications has delivered its first promises and that many novel contributions will be done over the next years. To better understand the field, the conceptual organization by different statistical methods presented here allows readers to easily put into context novel approaches to be published in the future.

REFERENCES

Adams, W. H. et al. (2003). Semantic indexing of multimedia content using visual, audio and text cues. *EURASIP Journal on Applied Signal Processing, 2*, 170–185.

Akutsu, M., Hamada, A., & Tonomura, Y. (1998). Video handling with music and speech detection. *IEEE Multimedia, 5*(3), 17–25.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM, 26*(11), 832–843.

Barnard, K., & Forsyth, D. A. (2001). Learning the semantics of words and pictures. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada.

Benitez, A. (2005). *Multimedia knowledge: Discovery, classification, browsing, and retrieval* [doctoral thesis]. New York: Columbia University.

Benitez, A. B., & Chang, S. F. (2002). Multimedia knowledge integration, summarization and evaluation. In *Proceedings of the International Workshop on Multimedia Data Mining in conjunction with the International Conference on Knowledge Discovery & Data Mining*, Alberta, Canada.

Benitez, A. B. et al. (2002). Semantics of multimedia in MPEG-7. In *Proceedings of the IEEE International Conference on Image Processing, Rochester*, NY.

Blei, D., & Jordan, M. (2003). Modeling annotated data. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada.

Brunelli, R., Mich, O., & Modena, C. M. (1999). A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation, 10*(2), 78–112.

Carneiro, G., & Vasconcelos, N. (2005). Formulating semantic image annotation as a supervised learning problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.

Duygulu, P., Barnard, K., de Freitas, N., & Forsyth, D. (2002). Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the European Conference on Computer Vision*, Copenhagen, Denmark.

Ekin, A., Tekalp, A. M., & Mehrotra, R. (2003). Automatic video analysis and summarization. *IEEE Transactions on Image Processing, 12*(7), 796–807.

Feng, S. L., Lavrenko, V., & Manmatha, R. (2004). Multiple Bernoulli relevance models for image and video annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Cambridge, UK.

Forsyth, D., & Ponce, J. (2003). *Computer vision: A modern approach*. Prentice Hall.

Harabagiu, S., et al. (2000). Falcon: Boosting knowledge for answer engines. In *Proceedings of the Text Retrieval Conference*, Gaithersburg, MD.

Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction*. Springer.

He, X., Zemel, R. S., & Carreira-Perpiñán, M. Á. (2004). Multiscale conditional random fields for image labeling. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Cambridge, UK.

Hofmann, T., & Puzicha, J. (1998). *Statistical models for co-occurrence data* (No. 1635 A. I. Memo). Massachusetts Institute of Technology.

Jeon, J., Lavrenko, V., & Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada.

Kumar, S., & Herbert, M. (2003a). Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proceedings of the IEEE International Conference on Computer Vision*, Nice, France.

Kumar, S., & Herbert, M. (2003b). Man-made structure detection in natural images using causal multiscale random field. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Madison, WI.

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning*, San Francisco.

Lavrenko, V., Manmatha, R., & Jeon, J. (2003). A model for learning the semantics of pictures. In *Proceedings of the Neural Information Processing System Conference*, Vancouver, Canada.

Leonardi, R., Migliotari, P., & Prandini, M. (2004). Semantic indexing of soccer audio-visual sequences: A multimodal approach based on controlled Markov chains. *IEEE Transactions on Circuits Systems and Video Technology, 14*(5), 634–643.

Li, B., & Sezan, I. (2003). Semantic sports video analysis: Approaches and new applications. In *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain.

Li, J., & Wang, J. Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(9), 1075–1088.

Lowe, D. (1999). Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision*, Kerkyra, Corfu, Greece.

Lu, L., Zhang, H-J., & Jiang, H. (2002). Content analysis for audio classification and segmentation. *IEEE Transactions on Speech and Audio Processing, 10*(7), 293–302.

Luo, Y., & Hwang, J. N. (2003). Video sequence modeling by dynamic Bayesian networks: A systematic approach from coarse-to-fine grains. In *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain.

Marr, D. (1983). *Vision*. San Francisco: W.H. Freeman.

Miller, G. A. (1995). Wordnet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41.

Mori, Y., Takahashi, H., & Oka, R. (1999). Image-to-word transformation based on dividing and vector quantizing images with words. In *Proceedings of the First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, Orlando, FL.

Naphade, M., et al. (1998). A high performance shot boundary detection algorithm using multiple cues. In *Proceedings of the IEEE International Conference on Image Processing*, Chicago.

Naphade, M., & Smith, J. (2003). Learning visual models of semantic concepts. In *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain.

Naphade, M. R., & Huang, T. S. (2000). *Stochastic modeling of soundtrack for efficient segmentation and indexing of video.* In *Proceedings of the Conference on SPIE, Storage and Retrieval for Media Databases*, San Jose, CA.

Naphade, M. R., & Huang, T. S. (2001). A probabilistic framework for semantic video indexing

filtering and retrieval. *IEEE Transactions on Multimedia, 3*(1), 141–151.

Natsev, A., Naphade, M., & Smith, J. (2003). Exploring semantic dependencies for scalable concept detection. In *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Los Angeles: Morgan Kaufmann Publishers.

Quattoni, A., Collins, M., & Darrell, T. (2004). Conditional random fields for object recognition. In *Proceedings of the Neural Information Processing Systems Conference*, Vancouver, Canada.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE, 77*(2), 257–286.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 888–905.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349–1380.

Snoek, C. G. M., & Worring, M. (2005). Multimedia event based video indexing using time intervals. *IEEE Transactions on Multimedia, 7*(4), 638-647.

Snoek, C. G. M., & Worring, M. (2005). Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications, 25*(1), 5–35.

Souvannavong, F., Merialdo, B., & Huet, B. (2003). Latent semantic indexing for video content modeling and analysis. In *Proceedings of the TREC Video Retrieval Evaluation Workshop*, Gaithersburg, MD.

Srikanth, M., Varner, J., Bowden, M., & Moldovan, D. (2005). Exploiting ontologies for automatic image annotation. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil.

Tan, Y-P., Saur, D. D., Kulkarni, S. R., & Ramadge, P. J. (2000). Rapid estimation of camera motion from compressed video with application to video annotation. *IEEE Transactions on Circuits and Systems for Video Technology, 10*(1), 133–146.

Tansley, R. (2000). *The multimedia thesaurus: Adding a semantic layer to multimedia information* [doctoral thesis]. University of Southampton, UK.

TRECVID. (2004). *TREC video retrieval evaluation*. Retrieved November 2005, from http://www-nlpir.nist.gov/projects/trecvid/

Tseng, B. L., Lin, C-Y., Naphade, M., Natsev, A., & Smith, J. (2003). Normalised classifier fusion for semantic visual concept detection. In *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293–302.

Vailaya, A., Figueiredo, M., Jain, A., & Zhang, H. (1999). A Bayesian framework for semantic classification of outdoor vacation images. In *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases VII*, San Jose, CA.

Vailaya, A., Figueiredo, M., Jain, A. K., & Zhang, H. J. (2001). Image classification for content-based indexing. *IEEE Transactions on Image Processing, 10*(1), 117–130.

Vasconcelos, N., & Lippman, A. (1998). A Bayesian framework for semantic content characterization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA.

Vasconcelos, N., & Lippman, A. (2000). A probabilistic architecture for content-based image retrieval. In *Proceedings of the IEEE Computer Vision and Pattern Recognition, Hilton Head*, SC.

Wang, Y., Liu, Z., & Huang, J-C. (2000). Multimedia content analysis using both audio and visual clues. *IEEE Signal Processing, 17*(6), 12–36.

Westerveld, T., & de Vries, A. P. (2003). Experimental result analysis for a generative probabilistic image retrieval model. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada.

Westerveld, T., de Vries, A. P., Ianeva, T., Boldareva, L., & Hiemstra, D. (2003). Combining information sources for video retrieval. In *Proceedings of the TREC Video Retrieval Evaluation Workshop*, Gaithersburg, MD.

Yavlinsky, A., Schofield, E., & Rüger, S. (2005). Automated image annotation using global features and robust nonparametric density estimation. In *Proceedings of the International Conference on Image and Video Retrieval*, Singapore.

This work was previously published in Semantic-Based Visual Information Retrieval, edited by Y.J. Zhang, pp. 334-354, copyright 2007 by IRM Press (an imprint of IGI Global).

Chapter 4.8
Principles of Educational Software Design

Vassilios Dagdilelis
University of Macedonia, Greece

ABSTRACT

Despite the generalized use of information and communication technologies (ICT) in teaching, their educational applications have not yet been standardized: a general consensus does not exist on how ICT can be applied to teaching nor on how educational software must be constructed. In this chapter, it is argued in favor of educational software construction being guided by a didactic problematique. In this framework we consider as a promising software category mindtools and, in particular, the so-called open microworlds. Their design must be guided by a number of principles: the tool logique, the multiple interface and the multiple representations principles. In this chapter, a detailed critique of these principles is also presented.

INTRODUCTION

From the time computers were invented until today, this latest decade has been intensely characterized, more than any of the previous ones, by the infiltration of the so-called new technologies in everyday life. Information and communication technologies (ICT) are also being integrated into education at all levels. In fact, this integration is a two-way process, as it has consequences for the educational system in which ICT are integrated. The influence of ICT on the way lessons are planned and conducted, on the administration of teaching institutions (schools, universities, and others), on existing teaching methodologies, on evaluation, and finally on the educational system in general, is so deep rooted, that ICT is likely to be the cause of a complete restructuring of the entire educational system.

Despite their use in teaching, which is constantly expanding, as well as the important influence they have on transforming current curricula, the educational applications of ICT have not yet been standardized, meaning that a general consensus does not exist on how ICT can be applied to teaching and a consensus does not exist that can be used as a general guideline for the development of educational software.

More specifically, concerning those points where a consensus of opinion does not exist, the following could be stated:

- **The characteristics of educational software:** There are many categories of educational software, which correspond to the different characteristics of that educational software as well as what its most appropriate use might be. The general characteristics of the software and its use are based on, explicitly or implicitly, learning theories, and pedagogic and didactic assumptions; therefore, these different points of view may be incompatible.
- **The usefulness of computers and educational software:** In certain situations, the effects of ICT on the educational system are obvious. For example, ICT have enabled long-distance learning to be reorganized based on new foundations due to the creation of computing networks and communication mechanisms among users at multiple levels (synchronous or asynchronous, with images, sound, video).
  In the majority of cases, even though the use of ICT is considered imperative, because, theoretically at least, ICT improve the lesson, this improvement, however, has not been amply documented. In other words, even though a basic reason for the use of ICT in education is the hypothesis that they improve both teaching and learning, the conditions that render teaching more effective are not actually known, and often, related research does not highlight any significant difference in the quality of the lessons based on ICT (http://teleeducation.nb.ca/nosignificantdifference/). Opinions have also been expressed which indirectly question the overall usefulness of ICT as a means of teaching support (Cuban, 2001; Stoll, 2000).
- **The validity and theoretical support of experimental data:** All over the world, experimental teaching methods and research studies are being carried out in order to discern the particular uses that will be most effective from a teaching perspective. One part of the research, as well as one section of the international literature related to the pedagogic uses of ICT, is geared at describing the characteristics of successful educational software. Unfortunately, research on teaching in the last few years, even though numerous, is not usually accompanied by a satisfactory theoretical background and thus has not managed to come up with substantial data to enable the design of adequate educational software and organize effective didactic situations.
- **The absence of a method for the construction of educational software:** Nowadays, two major categories appear to exist in the way educational software is constructed, each of which comprises several subcategories. In the first, the focus is on technology, meaning the technical features of the software and its construction method. In the second, the starting point is the teaching/learning process the software will support.

In this chapter, it is argued in favor of educational software construction being guided by a *didactic problématique*. Further, educational software must be created according to the most up-to-date learning theories and, more specifically, constructivism.

In this framework, the most promising software category is *mindtools* and, in particular, the so-called *open microworlds.*

The research to date, in conjunction with the large number of available educational software and, in general, digitalized educational material, have contributed to the relative progress that has taken place, but only in certain fields. Thus, concerning these mindtools, we have at our disposal a know-how that allows us to design, but only in certain situations, educational software and more general educational environments with notable functional features. This know-how can be sum-

marized in the form of general design principles on which an educational environment can be based (or at least some sections of it).

These principles are the following:

- The principle of *tool logic*: Computers and ICT in general, should be used as tools in order to make learning easier.
- The principle of multiple interface: The interface should offer the users the ability to express themselves not only by direct manipulation of objects but also with an active formulation of commands and instructions.
- The principle of multiple representations: Knowledge within the context of educational environments should be expressed in many ways, through multiple frameworks that are interconnected and equivalent from a functional point of view.

In the following paragraphs, a detailed critique of these principles is also presented.

EDUCATIONAL SOFTWARE DEVELOPMENT METHODS

Educational software was an area in which a medium level of production existed 20 years ago, however, in the last decade, in which time the graphical user interface (GUI) was established, production has become intense. Educational software originated mainly from two different sources, which in many cases cooperated: one source was software-producing companies and the other was the academic institutions, such as university laboratories or research centers and institutes (Harvey, 1995; van der Mast, 1995). This dual source essentially corresponds to the fact that in the last decade there has been significant progress, both in the level of technological know-how as well as in teaching theory, because current teaching know-how, despite the serious drawbacks pointed out in the introduction, has greatly improved in comparison to 20 years ago.

As was natural, the large amount of production gradually led to the development of various models of educational software, which can be used as guidelines in the creation of new educational software; a creation which, very likely may, in turn, lead to the improvement or even to a radical revision of existing models. Needless to say, this cycle can continue for a long period of time.

In the international literature, numerous models on the construction of educational software have been proposed, such as that described by Lage et al. (2002). This representative model proposes the use of "an incremental prototype life cycle" of 11 stages along general lines that must be followed for the construction of educational software. It is obvious that this model has been influenced by the general methodology of software engineering and attempts to adapt its general characteristics to educational software.

This type of methodology, as well as others like it, even though they might be considered effective from the point of view of developing software, is independent of both its *content* and its *didactic use*. This model, which is structured according to the principles of software engineering, does not actually focus on the most important aspect of educational software, namely, its teaching characteristics. Therefore, the models in this category are useful in describing the process for the development of educational software but have little value as tools for analyzing how useful particular educational software is for teaching purposes.

Another method used generally in the design of educational means and environments (Instructional Design) can also be used, such as the model ADDIE (Analysis, Design, Development, Implementation, Evaluation). This category is usually referred to as *tutorial* software.

One final category we often come across is that which proposes lists of practical advice or checklists covering certain aspects of educational

software (Beale & Sharples, 2002), with the drawback that it does not present a comprehensive set of principles on which this type of environment can be constructed.

The development of models or methods for the construction and use of educational software is, as a rule, either oriented toward the rationale of software engineering, or, consequently lacking in teaching parity, refers to the teaching characteristics of the software and is thus often fragmentary.

It is obvious that the construction of educational software should be based on some method, otherwise it is in danger of failing, of costing too much, or of being greatly delayed. This method, however, should be subject to the didactic rationale and not the other way round (Reusser, 1994). The basic characteristic of educational software should be its didactic efficiency rather than its technological supremacy, even though this, of course, is also desired.

KNOWLEDGE THEORIES AND DIDACTIC PROBLÉMATIQUE

Behaviorism and Constructivism

The many types of contemporary educational environments are certainly well-known: from the simplest drill-and-practice-type software to the more complex educational microworlds, a gamut of educational applications exists with a wide range of technical features, the interface, the audience they are geared to, and the goals aimed at. The wealth of environments is enormous, and their categorizations can be based on numerous criteria. It is generally accepted that the design of educational environments, whether digital or not, is guided, even at a subconscious level, by the pedagogic theories of the manufacturers. The choice of a theory is so important that most times it determines, directly or indirectly, the characteristics of the educational environment or even its type. Therefore, a large section of the educational applications with essential aims to convey knowledge and skills, many e-books for instance, belong to the behaviorist school (Collins, 1996).

Behaviorism and constructivism are two widely spread learning theories that are fundamentally different in their outlooks.

Behaviorism, which appears at an international level to be waning, assumes that learning is the result of a stimulus–response process. Learning is seen largely to be a passive procedure, where the subject attempts to adapt to the environment.

The constructivist approach, at polar ends with the behaviorist, is based on the hypothesis that subjects construct their own, personal knowledge through interaction with the environment; while learning essentially consists of a teaching methodology based on assimilating new information and then adapting the subject's mental structures in such a way as to be compatible with the new data. In the constructivist learning model, the subject, in attempting to explain or to examine questions, formulates hypotheses; searches for ways to verify or disprove these hypotheses; interacts with his or her environment (both material and human); redirects the results of experiences; and constantly reconstructs intellectual structures, the mental shapes, in such a way as to integrate them with the new data. Therefore, the educational environments of the constructivist paradigm differ greatly from those based on the behaviorism, because they favor active learning, cooperation between learners, investigation, formulation of hypotheses, as well as their verification or disproof. In other words, they promote actions and activities similar to scientific work, even though in reality, scientists' work is fundamentally different than that of students' work.

The Constructivist Approach

To a large extent, the various categories of educational software correspond to those of learning

theories, which they are based on. For this reason, there is no general consensus on educational software. On the contrary, these various categories coexist. As is natural within the framework of the various schools of thought, extended argumentation has been put forward from all sides.

The main argument for the preference of the constructivist approach and, subsequently, for cognitive tools, can be expressed by borrowing the characteristic metaphor of Resnick et al. (1996), who, arguing in favor of the "constructivist" environments, claim that all parents would prefer their children to learn to play the piano rather than learn to use a stereo. Even though it is much easier to use a stereo, the experience offered in learning to play a musical instrument as well as the likelihood of composing musical scores is incomparably richer than merely using even the most advanced electronic device.

Of course, the counterargument to this is that if someone wants to listen to high-quality music, 99 times out of 100, they will use a stereophonic system rather than attempt the arduous task of learning to play the piano. Moreover, it is almost certain that even those people who know how to play the piano own some kind of stereophonic system.

Nevertheless, a basic part of the authors' reasoning remains indisputable: the richest experiences are not gained by the simplest means, perhaps the opposite occurs—experiences are usually analogous to *the power of expression* the means offers the user. In this way, learning the piano decisively contributes to the creation and development of a relationship between the person and music that is rich in feelings. It must be admitted, however, that sometimes this relationship requires the use of devices that reproduce rather than produce sound, that is, the stereo, not the piano.

What then, if any, is the conclusion? We could say that the choice of the *means* is determined by the type of *use* one requires. This is also the case for educational software—its characteristics depend on the use for which it was made.

Mindtools

Throughout human history, technology has obviously contributed to the development of the human intellect and especially to that of learning. Pea (1985) defined *cognitive technology* as "any medium that helps transcend the limits of the mind, such as memory, in activities of thinking, learning and problem solving." Within this more general framework, certain computing environments, more precisely the environments and their didactic use, are usually characterized as *cognitive tools* (Collins, 2000) or even as *mindtools* (Jonassen, 2000), i.e., "computer-based tools and learning environments that have been adapted or developed a function as intellectual partners with the learner in order to engage and facilitate critical thinking and higher order learning" (p. 9). This category of educational environments presents certain significant didactic features.

There are numerous definitions and a relatively adequate amount of literature on cognitive tools or mindtools. According to Jonassen, Peck, and Wilson (1999), an educational environment should support knowledge construction, explorations, learning by doing, learning by conversing, intellectual partners that support learning by reflecting. On the basis of this analysis, therefore, it is possible to evaluate the various sections and functions of educational software, or more precisely, of an educational environment. Those sections and functions that serve or support the above-mentioned uses are thus more desirable. For example, the existence of a system of written communication between learners is a desirable characteristic if it is a prerequisite for the creation of communication and subsequently a type of social interaction between learners; this, however, is a characteristic of particular significance in the development of knowledge on the subject according to current didactic theories (Laborde, 1983). It should be emphasized that this criterion is too broad and, as a consequence, has only limited usability.

The choice of a specific group of software and the knowledge theory supporting it, unavoidably creates a generalization of design principles or, more precisely, a generalized criterion: if the end result is the creation of an educational software, of an educational environment intended to function as a cognitive tool, then its various sections and their functions must also be in accordance with this same rationale.

OPEN ENVIRONMENTS AND OPEN MICROWORLDS

Cognitive tools are not defined in any one way, and there are many types of software that fall into this category. The principles of their design have a general character, and although they cannot be applied to all types of educational environments, they are nevertheless useful as guides in the design of many such environments.

Cognitive tools, by their nature, are *open environments*. Those environments that do not have predetermined lessons or activities that the learner must follow but allow for the free development of activities within the framework of an area of knowledge can be considered as open environments.

There are many types of such environments, from Logo to the current ones such as the environments of so-called Dynamic Geometry (Bellemain, 1992) and others that are based on the synthesis of preexisting components (Rochelle et al., 1999). They have a number of characteristics in common that characterize them as mindtools and that have been "well-established" and, up to a point, have a theoretical basis.

Despite the fact that in recent years, open environments have been recognized as valuable learning environments, certain reservations have also been stated. Collins (2000) asked whether the environment should encourage guided learning or exploratory learning, and highlighted the fact that open environments actually allow the learner to "play around" with the software. He, himself, however, argued in favor of environments which in the beginning are rather guided until the learners gain a certain level of skill in order to be able to use them on their own and who gradually will be liberated. In reality, however, all educational software functions within the particular context of a *teaching situation*. As a rule, learners are required to solve a problem or to explore the various angles of a problem. So, it is unlikely that they "play around" (Hoyles & Noss, 1993).

A special category of software with a large teaching potential includes the so-called *open microworlds*, that is, the environments in which predetermined activities and lessons do not exist, but where learners can define new objects (programs, geometric shapes, functions, natural laws) as well as the relationships between them, in order to experiment (Hoyles & Noss, 1993; Bellemain, 1992; Tinker, 1993).

The developments of open environments, which favor experimental learning and problem solving, have proliferated in recent years. In these environments, a specified teaching scenario or route does not exist; the teaching material that accompanies these environments simply indicates the type of didactic situation.

The potential of these environments is of particular importance to teaching: the ability to choose to decrease the teaching noise (that is, undesirable side effects such as extremely long calculations that can totally overshadow the real objective of the lesson), the intelligent help (in conjunction with intelligent messages), and the programming by demonstration (Bellemain, 1992; Cypher, 1994).

Furthermore, these environments allow learners to express their ideas, even if erroneous, and apply them until they come to a dead-end or to an obviously wrong conclusion (Dagdilelis et al., 2003). If we accept the basic principles of constructivism (Balacheff, 1986; translation from French), these are necessary conditions in order "to provoke a knowledge confrontation and thus

develop learners' perceptions." Resnick et al. (1996) considered data that make the environment familiar to also be important. This familiarity does not aim at giving the learner motives but rather is necessary in order for the learner to be able to use a representational model in order to find the solution to the proposed problem. Such tools enable learners to represent their thought processes in external models for examination and reflection, and may further help them to improve these processes (Emrah, 1995).

The general principles described below are related to open environments designed with the rationale characteristic of the bulk of mindtools.

The Principle of *Tool Logic*

By nature, cognitive tools consist of environments in which the learner can explore phenomena from various angles and, by experimenting with these, reach certain conclusions. They are connected, either directly or indirectly, with possibly more than one field of knowledge (for example, working with Excel on a certain kind of problem requires a combination of programming and mathematics). They often provide the learner with the representation of a microworld, either natural or intellectual, and at the same time a series of tools needed for its exploration. For example, the so-called Dynamic Geometry environments [such as Cabri-Geometer (Bellemain, 1992) or Geometer Sketchpad (Roschelle & Jackiw, 2000)] consist of a simulation of Euclidian space (a totally mental construction that does not exist in reality), whereas the environments for the study of natural phenomena (such as Interactive Physics, 2003) represent natural space in which Newton's laws hold.

An important issue that arises is one concerning the dichotomy between an environment of "physical and epistemological faithfulness" (Collins, 2000). A true natural representation means that the educational environment depicts the actual situation, whereas an epistemological one means that in the educational environment, the same laws (mathematical, natural, etc.) apply, which also exist in nature. The most appropriate choice for either one or the other situation, or to be more precise, the necessity for natural faithfulness, because the other option is regarded as imperative, depends on the use the software is geared for. For example, the study of space from a geometric perspective might demand its simulation in order for the learner to be able to relate to and solve geometric problems under "real" conditions, such as measuring the distances of two points, which between them are not visible, or measuring the area of a golf course, which has an odd shape with a small lake in the middle. On the other hand, where the object of study might be the so-called theoretical (Euclidian) geometry, in which case the environment must "simulate" Euclidian space that, as stated above, does not actually exist because it is simply an intellectual construction, scientific faithfulness is adequate and sufficient.

The existence of natural faithfulness is a significant factor in the learning of concepts and relationships on account that it contains elements familiar to the learner. Furthermore, the existence of natural faithfulness makes it easier to apply the principle of *authenticity*, in other words, the teaching requirement for contextualized learning in an environment that is as close as possible to reality (Collins, 2000). Even though this principle has validity, the main object of research in a cognitive tool is not so much natural faithfulness as epistemological faithfulness, that is, the agreement of the environment and the reactions with the simulated system (Resnick et al., 1996).

Epistemological or physical faithfulness is not, however, a sufficient condition to justify the use of ICT. For example, an environment in which simply the construction of geometric shapes is possible does not make it a cognitive tool. The criterion is the principle of tool logic and its *didactic economy.* An educational environment contains a particular tool that supports a teaching method or the learn-

ing of certain concepts, and it should be designed as such. In addition, however, the potential of the teaching environment should be higher than that of usual environments. For example, software for physics should offer greater possibilities than that available in an actual laboratory. In other words, the design of an educational environment should be based on specific teaching needs, which are revealed by research on teaching or from the experience of teachers, rather than the other way around. This, of course, does not mean that the educational environment cannot break into new possibilities, but rather that in some way, the educational software should comprise the best possible means for the teaching of a concept or a technique.

Tool logic can, of course, vary as to its exact content from one area of knowledge to another and from one software program to another. Nevertheless, certain characteristics of educational environments, considered as cognitive tools for the learner, seem to play important roles in the learning process:

1. The environment should combine conviviality and usability. This means that the educational software should be simple and easy-to-use. An environment that requires complex procedures often forces the learner to focus on technical details instead of concentrate on the problem at hand. Novice programmers, for instance, often focus on dealing with the syntactic errors instead of with the construction of the algorithm—this is precisely the kind of *teaching noise* mentioned above. This fact has led to the creation of simpler languages in order to introduce learners to programming. A practical measure of usability is the effort and time needed to accomplish a task. More generally, the key criterion of a system's usability is the extent to which it enables people who use it to understand it, to learn and to make changes (Nielsen & Mack, 1994).
2. The possibility of avoiding *teaching noise* focuses on the essence (Tall, 1993). Teaching noise does not only come from a nonoptimal interface, as was the case in the above paragraph, but also from the use of methods and procedures that are inappropriate for the user. An essential function of educational environments is that in solving a problem, the user does not get caught up in any time-consuming processes that are counterproductive to learning. For example, a software such as Excel executes complex calculations with high precision and thus avoids teaching noise. This means that any in-between procedures that might take a lot of time and do not benefit learning are done away with. Usually, teaching noise can initially be an *object of learning*, which evolves into a means; and teaching economy often imposes its marginalization in the process of problem solving (Douady, 1993). For instance, basic calculations are an object of learning in the first years of schooling, which however, become the means to solving complex problems in physics and mathematics in the 11^{th} or 12^{th} year of school. Now, if a software allows the problems to be completed quickly, then this releases the learner from a tedious process that has no learning benefit (only the execution of many basic calculations). For this reason, in many software programs, the available information or representations can be increased or reduced, depending on the needs of the lesson (Bellemain, 1992; Dagdilelis et al., 2003).
3. The environment's large potential for expression is revealed in stages through the possibility of creating composite structures, with more specific and stronger tools. The procedures in Logo, the macroconstructions in the environments of dynamic geometry, and the formulae in Excel and the other spreadsheets, are examples of this type, because they offer the user the possibility to

compose new objects of the corresponding space (construction processes, geometric shapes, and algebraic relationships, respectively) based on the software's relatively simple primitives.

4. The possibilities to adapt to the user's needs and provide teaching help are given. This is perhaps the most difficult prerequisite from a technical point of view, because adapting to the needs of the user is materialized at multiple levels. For example, a module of the system can offer intelligent guidance in the use of the software. But, adaptation can be extended to more interesting teaching fields. For instance, in a contemporary educational environment, recordability, that is, the ability to continuously record the user's actions, is now a common option that presents interesting possibilities (Bellemain, 1992). Besides the obvious use of these records by the teacher, who, at times, with the help of the software itself, can analyze the learners' actions and gain valuable insights and results, the system can diagnose certain characteristics of the learners and offer them the necessary help by showing them the weaknesses in their solutions or even help them solve the problem in a variety of ways. In certain environments (such as Cabri-Geometer; Bellemain, 1992) of this type, the software can decide on, for the accuracy or not of a question the learner is formulating (e.g., are these points on the same straight line?), whether to point out the possible weaknesses of a program or to propose improvements or even correct solutions reached by the learner.
5. The possibility of communication is provided. Contemporary constructivist theories believe that humans do not learn in isolation but construct their knowledge through interacting with their environment, which consists of other people, including co-learners and teachers. The progress of network technology in the last decade has created the preconditions for the development of communication systems, even between learners and teachers who are geographically separated. However, it must be emphasized that, despite all its progress, technology is not yet adequately functional, and despite the theories that support the value of communication, we have but a few satisfactory examples of use that surpass the level of being commonplace.

Within tool logic, teaching economy imposes the integration into the environment of those elements that have a meaning from a teaching point of view. Therefore, the parts that characterize the environment, such as the multimedia elements, should serve some teaching goal and should not exist merely for effect.

The most important consequence of the principle of tool logic, however, is the need for educational software to be incorporated *into teaching.* This fact means that it is essential for educational software to be accompanied by a description of the corresponding didactic situations, within the framework in which the application will be used. This is necessary because even the best software is meaningless if it does not function within the context of a didactic situation (Brousseau, 1986a, 1986b). On the other hand, software exists that, although not educational, can be used as a cognitive tool within teaching, such as Excel and spreadsheets generally, as well as mathematics software such as Mathematica and the like.

The opposite phenomenon can, of course, be assumed, of the use of extremely rich software in an elementary way, especially in the situation where a corresponding didactic problématique does not exist.

ROLE OF THE INTERFACE AND THE PRINCIPLE OF MULTIPLE INTERFACE

The interface of contemporary educational environments plays an incredibly significant role in the process of learning, because it essentially determines the way in which learners will formulate their ideas. In this way, the interface indirectly determines an environment's expressive power (Climaco et al., 1993). The expressive power of an environment indicates how quickly and simply the tools enable a description of situations the users can immediately perceive in terms of existing goals and needs. It stands to reason that the environment should support a means of expression that will be simple but at the same time have the ability to formulate complex concepts. These two requirements can, however, be antithetical. In reality, there are at least two different systems that serve them.

Modern interface is characterized by the existence of graphics (GUI). In these new graphic environments the specific elements are the "*image-action*" and the *objectification* metaphor: actions are substituted by the duo *icons* and *actions with the mouse* (for example, clicking on the icon of a diskette has the result of saving the active file). This progress of the interface, in turn, leads to the development of a very interesting ability, that of direct manipulation (Bellemain & Dagdilelis, 1993). With direct manipulation, the user can "handle" objects (their representations) on the monitor, using a mouse directly on the image of the "object." An obvious characteristic of direct manipulation is, of course, that it does not require any kind of formulation on the part of the user. In order to destroy a file, all that is needed is to "throw it away" and to "empty" the "recycling bin." This, however, does not necessarily mean that the nonexplicitly expressed commands in every environment are direct use: in the *e-examples* of NCTM (National Council of Teachers of Mathematics), for instance, there is also a tiny programming environment for primary school children that does not require any written formulation—they use an iconic language—while in the past, there were other software with similar characteristics. In modern graphic environments, concepts and their relationships acquire a pseudomaterial basis, they appear to be—and behave as though they are—materials; for example, in the environments of Dynamic Geometry, the straight-line sections act as though they were elastic (they can elongate or shrink simply by dragging and dropping their ends), and in certain environments of Dynamic Algebra, the graphic representations function as though they are wire (they too permit modifications such as parallel transformation or flexure simply by drag and drop; Function Probe, 1999).

Direct manipulation, in this way, acquires particular importance as a teaching possibility, because it allows the user to express, in a direct way, relationships and choices that, in everyday terms, are unclear, because usually the user can *act* on them but not *express* them. In this way, direct manipulation gives the ability to choose a shade of color from a palette, the free-hand design of a shape, or the construction of a digital object, activities that are possibly familiar to the user but cannot easily be described (Eisenberg, 1996). Moreover, the conflict between "I do" and "I explain how to do" (Duchateau, 1992), is well-known as one of the common difficulties of novice programmers. For example, novice programmers can easily understand the design rule of a recursive shape, such as embedded squares or Koch's snowflakes, but they come up against great difficulty when they attempt to construct a recursive procedure that designs them (Carlisle, 2000; Dalit, 2001).

The advantage that direct manipulation has on teaching is particularly obvious in the cases where it is an essential *teaching variable*, that is, a factor that can greatly affect the teaching and learning processes by being present or absent. Early educational software for geometry, for in-

stance, which were mainly based on the explicit formulation of commands (such as GEOMLAND; Sendov & Sendova, 1995), turned out to be more difficult to use by young learners. In certain cases, the use of the "objects" or at least certain elements of the system is carried out through intermediate iconic mechanisms, such as sliders, of which Microworlds Pro and Avakeeo are well-known examples (Microworlds Pro, 1999; Koutlis & Hatzilacos, 1999). Nowadays, technology even allows for commands to be given indirectly to computers by demonstration (Bellemain, 1992; Cypher, 1994).

Direct manipulation comprises one end of the scale, while on the other end are the explicit formulation environments (such as Logo). Nevertheless, the inexplicit manipulation of objects is not the only solution and can coexist with the explicit expression of commands. International research shows (di Sessa et al., 1995; Hoyles, 1995; Sendov & Sendova, 1995; Laborde & Laborde, 1995) that direct manipulation is neither the only means nor the most appropriate solution for all situations. If direct manipulation offers ease in expression, explicit formulation offers the possibility to express composite concepts and relationships. Written formulation offers a large expression potential to the user, who can describe sophisticated relationships such as recursion or patterns with much greater ease than with either direct or indirect manipulation. In reality, the formulation of definitions, properties, and relationships is a component of the concepts themselves; it requires specialized knowledge that should be cultivated in the learner (Laborde, 1983).

The majority of typical languages, which enables a high degree of structure, are usually programming languages of some sort. Programming languages, of course, have their own teaching value but are, in some way, incredibly decontextualized, which means that they function in abstract and, thus, are often difficult to learn. On the contrary, in recent years, a generation of specific languages has appeared that has been adapted to the particularity of the environments in which they function. Thus, languages are being developed that allow the formulation of geometric or other causal dependencies (Sketchpad, 2003). In fact, the existence of these languages can make the software much more functional, because up to now, the tendency has been for software development to constantly add new features and tools, resulting in the creation of environments with scores of abilities but with problematic in-depth use (Eisenberg, 1996). It is worth noting that this problem as well as its suggested solution were highlighted 30 years ago in the area of programming. Dijkstra (1972) characterized certain languages as "baroque monstrosities" and argued in favor of languages with a small repertory of commands but that had a high possibility of composing new ones.

Particular teaching problems, such as the choice of functional syntax and the appropriate formalism arise due to the existence of specific descriptive languages. However, with the use of various techniques, these problems can nowadays be dealt with to a satisfactory degree.

So, as current research shows, each of the methods of expression (manipulation or formulation) has its advantages and disadvantages; therefore, the best construction strategy for educational environments is to incorporate both methods that have the capacity to be used for contemporary needs. Perhaps the best examples are computers and their operating systems. The two most popular systems (Windows and Unix in various forms) coexist, serving different groups of users and converging to become a "double" nature, as it were—Windows has written commands, while for the Unix-like systems, X-Windows have now become an important component.

MULTIPLE REPRESENTATIONS

The progress of technology (high-analysis monitors, fast processors and graphic cards, effective

algorithms, etc.) has enabled, to a great extent, the development of technical illustrations with animated or unanimated images or video. Current educational environments are using these new means all the more extensively: animated images that are also interactive. The particularly significant role of images has often been emphasized in the international literature (Tall, 1993; Kaylan, 1993).

The progress of technology has also increased the potential of another section of educational environments, namely, that of *multiple representations*. A representation, in this case, is a system that symbolizes certain knowledge. For example, a mathematical function can be represented as a formula or as a graph; an algorithm can be represented in the form of diagrams or in a textual programming language.

Support multiple representational forms enable the learner to represent concepts and meaningful relationships between concepts both verbally and pictorially (Kozma, 1992) and allow the learner access to the different representations simultaneously, so that interrelationships are directly available (Emrah, 1995).

Multiple representations are not necessarily totally based on pictures. According to Beale et al. (2002), they include interactive systems of "external representations. External representations are observable, and often manipulatable, structures, such as graphs, tables, diagrams or sketches that can aid problem solving or learning" (p. 18). Here, external representations are regarded in the broad sense of the term, including in them *textual representations*, in other words, subordinate texts. The importance of external representations is that each one can comprise a significant teaching aid for the learner, because it contains important information. For example, these types of representations can make obvious the relationship between data-producing sets with common characteristics, tendencies, or properties. Of course, these representations have the disadvantage that, although they make certain properties of the concepts under study obvious, they leave out certain others. Often, one type of external representation is particularly useful in order to solve a certain category of problems but is inappropriate for all the others. This drawback is overcome by the existence of multiple external representations.

Of particular interest are those external representations that are dynamic and interconnected and in fact bidirectional. In other words, they present *symmetrical possibility*. In educational environments, one form of expressing a concept or relationship is often preferred over another, due to the particular *didactic contract* (Brousseau, 1986) that commonly exists in school environments or simply due to technical reasons. Thus, a graphic representation of a function can often result from its analytical expression, and the change to the analytical expression causes a modification of the curve but not the opposite. This modification has been previously pointed out (Bellemain & Dagdilelis, 1993; Schwarz, 1993) and integrated into earlier pilot software or more recently, into well-known software (such as Function Probe, 1999). However, this possibility is limited, as it can be used only for certain categories of functions that are known beforehand. At school, most explored functions are known beforehand, so this is not a concern.

Recent research studies on teaching in various scientific fields have shown (Douady, 1993; Tall, 1993) that *multiple representations* clearly have teaching value. For example, the representation of a mathematical function in an analytical, graphical, and numerical way with the ability of direct interaction between the various frameworks of expression allows the learner to produce more complete images of the concept being examined and to develop significant notions on it, even when lacking basic knowledge (Tall, 1993). Environments should offer the option of choice between one or multiple representations, depending on the teaching needs at the time. In this way, for instance, multiple representations for computing

programs executed simultaneously on DELYS (Dagdilelis et al., 2003)—a software developed by the University of Macedonia (Thessaloniki, Greece) for the teaching of programming—the user can decide which functions will be visible and which will not.

SYNTHESIS

In the above paragraphs, three basic principles for the design of modern educational environments were presented: the principle of tool logic, the principle of multiple interfaces, and the principle of multiple representations. Although the number of such principles has not been exhausted, as the research data show, they contribute significantly to the development of educational environments with high didactic specifications.

The rationale behind the examination of these design principles was focused on their teaching functions rather than on the logic of software engineering. This is the reason why general principles related to those categories of educational software that appear to be the most promising, that is, cognitive tools, were chosen. The main conclusion from the study of these educational environments is that a basic factor for their effective use in teaching is the *didactic situation in which they are applied*. The value of cognitive tools lies precisely in their ability to support situations with a rich teaching potential. With few exceptions, research on teaching does not seem to have progressed at the same rate as that on the technology of educational software. This fact explains the general impression, which exists internationally, that educational software has not yet succeeded in achieving its goal to improve teaching to the level expected. This is a direction that should be followed in order to succeed in the design and construction of even better quality educational environments and also simultaneously to make better use of existing software.

REFERENCES

Balacheff, N. (1986). *Une étude des processus de preuve en mathématique chez les élèves de collèg*, Thèse de Doctorat d'Etat en Didactique des mathématiques, Grenoble, France.

Beale, R., & Sharples, M. (2002). *Design guides for developers of educational software*. British Educational, Communication and Technology Agency.

Bellemain, F. (1992). Conception, realisation et experimentation d'un logiciel d'aide à l'enseignement lors de l'utilisation de l'ordinateur. *Educational Studies in Mathematics*, 23, 59–97.

Bellemain, F., & Dagdilelis, V. (1993). *La manipulation directe comme invariant des micromondes educatifs.* Fourth European Logo Conference, Athens, Greece.

Brousseau, G. (1986a). Fondement et methodes de la didactiques de mathématiques. *Recherches en Didactique des Mathématiques, 7*(2).

Brousseau, G. (1986b). *Theorisation des phénomènes d'enseignement des mathématiques.* Thèse d'etat, Univesité de Bordeaux I, 1986.

Carlisle, E. G. (2000). Experiences with novices: The importance of graphical representations in supporting mental models. In A. F. Blackwell, & E. Bilotta (Eds.), *Proc. PPIG* 12 (pp. 33–44).

Climac, J. N., Antunes, C. H., and Costa, J. P. (1993). Teaching operations research using "home made" software. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science*, NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 305–338). Springer Verlag.

Collins, A. (1996). *Design issues for learning environments, international perspectives on the design of technology-supported learning environments.* S. Vosniadou, E. De Korte, R. Glaser, &

H. Mandl (Eds.). Mahweh, NJ: Lawrence Erlbaum Publishers.

Cuban, L. (2001). *Oversold and underused: Computers in classrooms.* Harvard, MA: Harvard University Press.

Cypher, A. (Ed.). (1994). *Watch what I do—Programming by demonstration.* Cambridge, MA: The MIT Press.

Dagdilelis, V., Evangelidis, G., Satratzemi, M., Efopoulos, V., & Zagouras, C. (2003). DELYS: A novel microworld-based educational software for teaching Computer Science subjects. *Computers and Education, 40*(4).

Dalit, L. (2001). Insights and conficts in discussing recursion: A case study. *Computer Science Education, 11*(4), 305–322.

di Sessa, A., Hoyles, C., Noss, R., & Edwards, L. (1995). Computers and exploratory learning, setting the scene. In di Sessa et al. (Eds.). *Computers and exploratory learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 1–12). Springer Verlag.

Dijkstra, E. W. (1972). The humble programmer. *Communications of the ACM, 15*(15), October, 859–866.

Douady, R. (1993). L'ingenierie didactique, un moyen pour l'enseignant d'organiser les rapports entre l'ensegnement et l'apprentissage. *Cahier de DIDIREM, 7*(19).

Duchateau, C. (1992). From DOING IT…to HAVING IT DONE BY... The Heart of Programming. Some Didactical Thoughts, NATO Advanced Research Workshop *Cognitive Models and Intelligent Environments for Learning Programming,* S. Margherita, Italy.

Eisenberg, M. (1995). Creating software applications for children: Some thoughts about design. In A. A. diSessa, C. Hoyles, & E. Noss (Eds.), *Computers and exploratory learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 146. Springer Verlag.

Emrah, O. (1995). Design of computer-based cognitive tools. In A. A. diSessa, C. Hoyles, & E. Noss (Eds.), *Computers and exploratory learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 146. Springer Verlag.

Function Probe. (1999). Retrieved from the World Wide Web: http://questmsm.home.texas.net/

Harvey, J. (1995). *The market for educational software, Critical Technologies Institut- RAND,* prepared for Office of Educational Technology, U.S. Department of Education. DRU-1 04 –CTI.

Hoyles, C. (1995). Thematic chapter: Exploratory software, Exploratory cultures? In di Sessa et al. (Eds.), *Computers and exploratory learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 19–219). Springer Verlag.

Hoyles, C., & Noss, R. (1993). Deconstructing microworlds. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 385–413). Springer Verlag.

Interactive Physics. (2003). Retrieved from the World Wide Web: http://www.interactivephysics.com/description.html

Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking.* New Jersey: Merrill/Prentice Hall.

Jonassen, D. H., Peck, K. C., & Wilson, B. G. (1999). *Learning with technology: A constructivist perspective.* New Jersey: Merrill/Prentice Hall.

Kaylan, A. R. (1993). Productivity tools as an integrated modeling and problem solving environment. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science,* NATO ASI Series, F: Computer and

System Sciences, Vol. 146 (pp. 439–468). Springer Verlag.

Koutlis, M., & Hatzilacos. (1999). *"Avakeeo": The construction kit of computerised microworlds for teaching and learning Geography.* Retrieved from the World Wide Web:= http://www.ncgia.ucsb.edu/conf/gishe96/program/koutlis.html

Kozma, R. B. (1992). Constructing knowledge with learning tools. In P. A. M. Kommers et al. (Eds.), *Cognitive tools for learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 81 (pp. 305–319). Springer Verlag.

Laborde, C. (1983). Langue naturelle et écriture symbolique, deux codes en interaction dans l'enseignement mathématique. Didactique des mathématiques.

Laborde, C., & Laborde, J. M. (1995). What about a learning environment where Euclidean concepts are manipulated with a mouse? In di Sessa et al. (Eds.), *Computers and exploratory learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 241–262). Springer Verlag.

Lage, F. J., Zubenko, Y., & Cataldi, Z. (2001). An extended methodology for educational software design: Some critical points. *31th ASEE/IEEE Frontiers in Education Conference, T2G-13, 2001,* Reno, Nevada, USA.

MicroWorlds Pro. (1999). *Logo Update On Line*, Vol. 7, Number 2.

Nielsen, J., & Mack, R. (1994). *Usability inspection methods.* New York: John Wiley & Sons.

Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist, 20*(4), 167–182.

Resnick, M., Bruckman, A., & Martin, F. (1996). Pianos not stereos: Creating computational construction kits. *Instructions, 3*(6).

Reusser, K. (1994). Tutoring mathematical text problems: From cognitive task analysis to didactic tools. In S. Vosniadou, E. De Corte, & H. Mandl (Eds.), *Technology-based learning environments,* NATO ASI Series, F: Computer and System Sciences, Vol. 137 (pp. 174–182). Springer Verlag.

Roschelle, J., & Jackiw, N. (2000). Technology design as educational research: Interweaving imagination, inquiry & impact. In A. Kelly, & R. Lesh (Eds.), *Research design in mathematics & science education* (pp. 777–797), Mahwah, NJ: Lawrence Erlbaum Associates.

Roschelle, J., DiGiano, C., Koutlis, M., Repenning, A., Jackiw, N., & Suthers, D. (1999). Developing educational software components. *IEEE Computer, 32*(9).

Schwarz, J. L. (1993). Software to think with: The case of algebra. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 469–496). Springer Verlag.

Sendov, B., & Sendova, E. (1995). East or West—GEOMLAND is best, or Does the answer depend on the angle? In di Sessa et al. (Eds.), *Computers and exploratory learning,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 59–78). Springer Verlag.

Sketchpad. (2003). Retrieved from the World Wide Web: http://www. keypress.com/sketchpad/

Stoll, C. (2000). *High tech heretic: Reflections of a computer contrarian.* Anchor Books.

Tall, D. (1993). Interrelationships between mind and computer: Processes, images, symbols. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science,* NATO ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 385–413). Springer Verlag.

Tinker, R. F. (1993). Modelling and theory building: Technology in support of student theorizing. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science,* NATO

ASI Series, F: Computer and System Sciences, Vol. 146 (pp. 91–114). Springer Verlag.

Van der Mast, C. (1995). Developing educational software: Integrating disciplines and media (pp. 1–96). Ph.D. thesis, Technische Universiteit Delft.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, p. 113-134, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.9
Online Multimedia Educational Application for Teaching Multimedia Contents:
An Experiment with Students in Higher Education

Alcina Prata
Higher School of Management Sciences (ESCE), Portugal

Pedro Faria Lopes
Higher Institute of Labour and Business Studies (ISCTE), Portugal

ABSTRACT

This chapter describes an experiment undertaken with higher education level students, which consists of utilizing an online multimedia educational application as an aid in teaching organizational multimedia. This course is taught to fourth year students at the Escola Superior de Ciências Empresariais (Higher School of Management Sciences, Setúbal, Portugal), where the first author teaches. This chapter also describes the educational software model used for the planning, development and evaluation of the above-mentioned application. This model is the result of the integration of the model presented in the first author's Master's thesis with methods, methodologies and guidelines proposed by others. As for the resulting application, the manner in which it was applied and its evaluation are also presented in this document. The results obtained are then interpreted and future developments proposed.

INTRODUCTION

The use of information technology currently plays an important part in the day-to-day of the majority of public and private institutions. The traditional educational system also has had to adapt to this new way of doing things (Chambel et al., 1998) and has benefited significantly from the contribution of these types of technological applications (Azevedo, 1997; Hartley, 1999; McCarthy, 1995). Likewise, the reigning "professional philosophy"

has also evolved towards the notion of life-long learning (Ryan et al., 2000). Professional careers are becoming increasingly demanding, implying a rapid adaptation to new circumstances and constant acceleration in education, preferably without dismissing employees (Abbey, 2000; Chute et al., 1999). All this leads us to e-learning: a teaching method which utilizes Internet technologies to supply, at a distance, a range of solutions for the acquisition and/or updating of knowledge (Machado, 2001; Rosenberg, 2001; Ryan et al., 2000).

The main reason for the growing popularity of this teaching method is the fact that it combines the advantages of using information technology in education (Azevedo, 1997) with the advantages of distance learning (Machado, 2001; Rosenberg, 2001), namely, access to information using the new instructional model "anytime, anyplace and anybody" (Aggarwal, 2000). This was, therefore, the main reason for choosing to develop and use an Online Multimedia Educational Application (OMEA) to serve as an aid in the teaching of *Organizational Multimedia*, a course taken by fourth year students at the Higher School of Management Sciences. As this is a compulsory course, the classes tended to be very heterogeneous, bringing together students from five different academic areas. Given that the students had different schedules and study constraints, this naturally implied a few difficulties when teaching the course. So, with OMEA, the main goal was to solve this problem, and render access to information easier for anybody, anytime and anyplace. At the same time, we were expecting to benefit from the advantages of using information technology (namely Web-based technology) as a teaching support inside and outside classroom and, finally, to contribute to the modernization of study methods used by students. Other goals were to acquire more experience in planning, developing and evaluating online educational applications and to research the critical process of developing real educational, multimedia and interactive online applications. Unfortunately, in spite of all the advantages associated with the use of information technology (namely Web-based technology) as teaching support, we still have a long way to go in the planning and development of real educational, multimedia and interactive online applications. As a recent and supposedly lucrative phenomenon, e-learning is now concentrating all the efforts of the majority of teaching institutions (Palloff et al., 2001). However, this effort is mainly concentrated on the search for the "ideal" platform, instead of the "ideal" content. So, it is usual to find technologically strong platforms supporting very poor content. In a study to evaluate the online educational applications delivered by some university e-learning platforms (Prata et al., 2003a), we saw that the majority:

- Are not interactive
- Are mainly text-based
- Are not planned and developed by following adequate educational software models and educational and pedagogical rules
- Are made available, at least initially, without being properly tested, namely by getting some feedback from the target population

In the development process of the OMEA, which was supposed to serve as an aid in the teaching of the *Organizational Multimedia* course, we tried to overcome the problems mentioned above, namely by creating an online application:

- That could act as a substitute for the face to face classes that the students could not attend
- That would allow the students to achieve the same results they would achieve by attending the face to face classes
- That would correspond to the target population's expectations and needs
- That would be motivating
- As well as interactive
- And pedagogically adequate

- Which would include different types of media components, namely, text, image, graphics, sound, animation and video
- Which would follow an adequate educational software model for its planning, development and evaluation
- That would be appropriately tested
- That could be integrated as an help/support tool in the classroom
- That could be used by the students as a regular tool/support to their study
- That could be available on the Web on a short-term basis, and later easily adapted to the school-specific intranet and e-learning platforms
- That could be adapted to any standardized e-learning platform (LMS)

In the following sections of this chapter we will justify in more detail the need for this kind of study and its contribution to the resulting OMEA. We will also explain the characteristics of the OMEA developed, present the educational software model used for the planning, development and evaluation of this OMEA and explain the evaluation method used to assess it. Finally, the results obtained will be presented and interpreted, and future developments will be proposed.

BACKGROUND

In recent years information technology has become more commonplace in many areas partly due to government initiatives such as society digitalization, and due to the general public's growing awareness of the Internet as a privileged vehicle for obtaining information. The utilization of these technologies (online and off-line multimedia contents) is presently part of the day-to-day of the vast majority of public and private institutions. Traditional education systems also felt the need to adapt to this new society (Chambel et al., 1998) and have benefited significantly from the contributions of these types of contents (Azevedo, 1997; Hartley, 1999; McCarthy, 1995). Some studies made in order to evaluate the impact, in general terms, of using computers and multimedia in teaching, demonstrate that these types of environments promote creative rationalization, problem solving approaches, strategy formulation, and persistence in the pursuit of goals. Since they promote the use of different sensorial channels, they also represent substantial gains in terms of learning, retention and understanding of issues (Azevedo, 1997; McCarthy, 1995). In other words, they are a more stimulating and involving study method than traditional materials since they imply increased adaptability to different styles of learning, greater involvement of students in the learning process and, also, offer equal advantages for students with or without previous knowledge of information technology. In short, computers and multimedia are a consistent teaching method and, on a worldwide scale, a new learning model (Könyves-Tóth et al., 1995).

Another issue widely discussed nowadays, although not a recent phenomenon, is Long Distance Learning (LDL), which is characterized as an educational event where learning is undertaken by a physical separation (geographical and/or temporal) between students and teachers (Santos, 2000). LDL has emerged as a way of bringing flexibility to educational resources and of leaving less mobile populations the option of continuing their studies/further education. Increasingly, with time, that less mobile population has grown. The "professional philosophy" has also evolved towards a notion of life-long learning (Ryan et al., 2000), and professional careers in most institutions are becoming increasingly demanding, requiring rapid adaptation to new circumstances and constant education, preferably without employee absence to achieve these objectives (Abbey, 2000; Chute et al., 1999). LDL has become the only alternative for many. This teaching method has also evolved with time. What first started as an educational project by means of the post, or tele-school in the

1960s (classes broadcast over the traditional TV network) and CD-ROM based media in the 90s, is now often based on e-learning systems (which have been continuously refined since the Internet's appearance about 10 years ago) (Machado, 2001). E-learning can be defined as a teaching method which utilizes Internet technology to supply a set of solutions at a distance (Online Multimedia Educational Applications) for the acquisition and/or updating of knowledge, and there are many authors involved in its refinement (Machado, 2001; Rosenberg, 2001; Ryan et al., 2000).

According to Elliot Masie, one of the most respected and recognized specialists in this area, the excitement and commitment with which some countries, such as Portugal, embrace e-learning will make them progress at the speed of light (Machado, 2001). Steve Ryan believes that the development of e-learning, which is occurring in all types and at all levels of educational organizations (public and private), all around the world, is highly significant (Ryan et al., 2000). Whereas Marc Rosenberg considers that nowadays, almost all traditional American institutions for higher education are developing e-learning systems (Rosenberg, 2001), such systems are not limited only to higher education because, according to LeBaron, Laurel Springs High School was the first of its kind to implement all of its courses online (LeBaron, 2001).

The main reason for the growing popularity of this teaching method is the fact that it combines the advantages of using information technology in education (Azevedo, 1997) with the advantages of distance learning (Machado, 2001; Rosenberg, 2001); namely, access to information using the new instructional model "anytime, anyplace and anybody" (Aggarwal, 2000). These advantages, in relation to traditional study methods, are on the level of accessibility. Information:

- Is accessible to anybody at any time, anywhere (Aggarwall, 2000; Machado, 2001; Rosenberg, 2001)
- Is accessible through multimedia contents (Machado, 2001)
- Is ready to evolve at the student's individual rate (Machado, 2001)
- Uses hypertext (Machado, 2001)
- Focuses on the student – who is an active participant (Machado, 2001)
- Is available in module form (Machado, 2001)
- Relies on a flexible electronic infrastructure (Machado, 2001)
- Allows for simple and rapid updating (Machado, 2001; Rosenberg, 2001)
- Allows for a great diversity of operators (Machado, 2001)
- Works on individual programs (Machado, 2001)

The general advantages also pointed out are:

- Low cost (Machado, 2001; Rosenberg, 2001)
- Efficient proximity to an unlimited number of people (Aggarwall, 2000; Machado, 2001; Rosenberg, 2001)
- Feasibility of personalization (Machado, 2001; Rosenberg, 2001)
- Permanent availability (Machado, 2001)
- Ease of use. Does not require too much previous knowledge (Machado, 2001; Rosenberg, 2001)
- Universality, since it is based on Internet protocols which operate under standard protocols (Machado, 2001; Rosenberg, 2001)
- Promotion of the emergence of communities with common interests, functioning as a motivating factor (Machado, 2001; Rosenberg, 2001)
- Scalability, as it always allows inclusion of another person (Machado, 2001; Rosenberg, 2001)
- Promotion of a better collaborative environment (Machado, 2001)

- Justification and maximization of investments in intranet and Internet networks (Rosenberg, 2001)

In short, summarizing all that has been mentioned previously, and considering:

- All the advantages of utilizing multimedia and information technology in teaching
- The increasing importance of courses delivered through LDL
- The specific advantages of using e-learning systems
- The testimony of a wide range of known authors

… we can easily conclude that e-learning is here to stay and that it constitutes, on a worldwide scale, a new way of teaching that is redefining the concept of learning as we know it. That is, we are in the presence of an issue that offers great research prospects.

However, it is important to understand that technology is a tool and not a means in itself. It is absolutely vital that, in conjunction with technological investments, efforts be made to find methodologies, rules, guidelines and educational principles that are appropriate to the planning, development and evaluation of contents of this new way of teaching. The non-observance of this basic rule may lead to the development of systems that are technologically perfect but unable to fulfill their role from an educational point of view. This situation, for example, occurred in Portugal, when in the mid-nineties, companies concerned only with filling market niches started mass producing CD-ROMs, many of which were of low quality from a pedagogical point of view (Prata, 2000). In relation to e-learning, that imbalance is visible. It is noticeable that, with the majority of institutions (companies and schools, public and private) that have begun developing or are currently using e-learning solutions, the dominant concern has been essentially technological. That is, they have focused on developing the ideal platform in detriment of the educational slope, which should always be considered first when planning and developing contents for education.

One of the largest difficulties, despite more than four decades of research in human-computer interface areas, is that multimedia system architectures are still missing essential rules and guidelines for the association of different multimedia components. The use of contents with several types of multimedia components, such as text, graphics, image, sound, animation and video, is normally very attractive to users and helps retain their attention and interest during long periods of time. However, this is not the fundamental issue. What truly matters is to understand the real impact, in terms of efficiency and efficacy, these contents will have at the level of information processing and in the acquisition of knowledge. To achieve the desired efficiency and efficacy it is necessary, whilst developing the content, to consider the following factors:

- The way human beings learn
- The personal and cultural characteristics of the target population. Personal, amongst other things, in terms of age, education, previous knowledge, desires and expectations. Cultural, in terms of "certain cultural and policy cross-border peculiarities" (LeBaron, 2000)
- The specific characteristics of each component used (which, obviously, impedes to group or associate components randomly) (Chambel et al., 1998)
- The advantages and disadvantages of each component (for instance, video is, amongst all components, the most powerful in generating attitudes and emotions) (Guimarães et al., 2001)
- The specific characteristics inherent in the issues being presented (e.g., not appropriate to use the same methodology to explain such

different issues as literature and information technologies)
- The need to accommodate several different styles of learning (Chambel et al., 1998)
- The importance of interactivity and of active participation by the user (Chambel et al., 1998)
- The need for a virtual environment (learning environment) which facilitates the learning process. A learning environment is an active environment that infuses the user with a sense of mission that leads him or her to participate actively and to do things. It is an interactive environment (Abbey, 2000; Chambel et al., 1998).
- The learning process is a continuous one and not a set of sporadic and disconnected events (Machado, 2001).

In recent years, due to the growing success of the Internet, some authors have developed studies of generic rules/guidelines for the development of educational multimedia contents for the Web. Several universities have multimedia laboratories where these studies are made. Some of the most important are: the laboratories of the University of Alberta in Canada (Driedger, 1999), the University of Toronto (Drenner, 1998), Yale University (Lynch et al., 1999), the British Open University (OU) in England (Santos, 2000) and Universidad Nacional de Educación a Distancia (UNED) in Spain (Santos, 2000). However, the rules/guidelines defined by these institutions for the development of multimedia environments are few, generic, highly varied and constantly changing. From a study to compare different Web style guides it was possible to conclude that "sometimes they make quite similar recommendations for developing a web site, sometimes they disagree, and sometimes they emphasize different design considerations" (Berk, 1996). Specific rules which are defined and accepted world-wide do not even exist yet (DISA, 1995), which means that some empirical research is needed in order to determine/identify which design criteria will facilitate different online tasks.

This considerably aggravates some considerations/concerns previously pointed out as fundamental to the efficiency and efficacy of content development. These regard the selection of multimedia components to be used but more importantly the manner in which these components can be combined in user-friendly graphical interfaces and, simultaneously, be efficient from an educational point of view. The size and the implications of this deficit allow us to conclude that this is an area where a lot of development is needed and where any advance will be heartily welcome.

With the development of the OMEA our main goal was to facilitate access to information for anybody at anytime and anyplace; that is, to solve the students' problems in attending the face to face classes. Simultaneously, we were expecting to benefit from the advantages of using Web-based technologies as a teaching support inside and outside the classroom and to contribute to the modernization of the study methods used by students. Another goal was to acquire more experience in planning, developing and evaluating online educational applications. Taking into account the above-mentioned difficulties, we also committed to doing some research into the complex process of developing real educational, multimedia and interactive online applications.

ONLINE MULTIMEDIA EDUCATIONAL APPLICATION (OMEA)

As mentioned before, the OMEA main goal is to solve the students' problems in attending the face to face classes by facilitating access to information for anybody at anytime and anyplace. Thus, students who could not attend the *Organizational Multimedia* face to face classes were considered to be the target group for the OMEA and the most

important factor was to develop an OMEA that could best compensate for a student's absence. However, and since we also intended to use the application as a study support inside and outside the classroom, the application was developed so that the general student population could also use it.

All classes in the course are laboratory-based and last three hours. The first hour and a half is dedicated to theory (theory-based) and the remainder to practice (practice-based), that is, to the presentation of practical cases. Given that the educational software model used for the planning, development and evaluation of the OMEA relies on the initial development of a prototype (which, if proven efficient, will serve as the basis for the subsequent development of the final educational application), the OMEA will be, from this point forward, designated as a prototype.

However, the extension/diversity of the course content, which comprises classes covering six different themes, namely text, graphics, image, sound, animation and video, also had to be considered. There are lots of theories and experiments about learning at a distance, especially if we consider research conducted by the British Open University over the last 30 years (Santos, 2000). However, there are some difficulties inherent to the development of this type of OMEA. A universal "formula" capable of guaranteeing the success of applications does not exist and, in order to be properly presented, each theme has its own specifications and methodologies, especially with regards to practice-based classes where the differences appear the greatest. These were the reasons that led us to opt for the development of more than one prototype. As the approach to some of the above mentioned themes was similar, the decision was taken to implement four prototypes in the following order: the first with a practice-based class on image (Prata, 2003), the second with a practice-based class on sound (Prata et al., 2003b), the third with a practice-based class on animation (Prata et al., 2003d) and the fourth with a practice-based class on video (Prata et al., 2003c).

The prototypes were not developed simultaneously because:

- We wanted to learn from the different experiments; therefore a new prototype only began to be developed after the previous one had been concluded and tested
- From a technological point of view, the prototypes include components with different levels of complexity. In terms of manipulation, representation, storage, and types of resources needed, components may be classified in the following ascending level of complexity: image, sound, animation and video. Therefore, we decided to implement the prototype with a class on image first, and the prototype with a class on video last. This way we could count on some experience gained from our previous work. It is important to say that this last prototype was doubly challenging given that video is the most complex component, and although a great deal has been achieved in terms of improving compression algorithms and increasing bandwidth/access speeds to the Internet, we are still far from achieving the ideal
- We wanted to include as many groups of students as possible in the evaluation process. In fact, each prototype was evaluated by a different group of students. We paid particular attention to this question because we did not always want to bother the same students and we wanted each prototype to be evaluated by a new group that had never seen the previous prototypes. All the prototypes were evaluated by students taking the *Organizational Multimedia* course during the 2002/2003 academic year, both semesters.

Each of the four prototypes comprises three different sections: one section with the content of a practice-based class about a specific theme (image, sound, animation or video), the Frequently Asked Questions section (FAQs) and the exercise section (which includes solutions).

In conclusion, it is important to realize that the main goal of our work is to develop a final OMEA, which will include all the classes on the *Organizational Multimedia* course. However, as supported by the model we have used, we must start by developing and testing a prototype. As mentioned before, the content of the *Organizational Multimedia* course comprises six different chapters and our first thought was to develop a prototype for each chapter. Yet, after a detailed analysis we came to realize that some chapters included components with similar characteristics, so the final decision was to develop four prototypes: a first prototype with a practice-based class on image (Prata, 2003), a second one with a practice-based class on sound (Prata et al., 2003b), a third one with a practice-based class on animation (Prata et al., 2003d), and finally, a fourth one with a practice-based class on video (Prata et al., 2003c).

As part of a major project, the educational software model used for the planning, development and evaluation of the prototypes as well as the evaluation method used were the same in all four prototypes. Both the educational software model and the evaluation method used will be described in the next sections.

EDUCATIONAL SOFTWARE MODEL

Origins of this Model

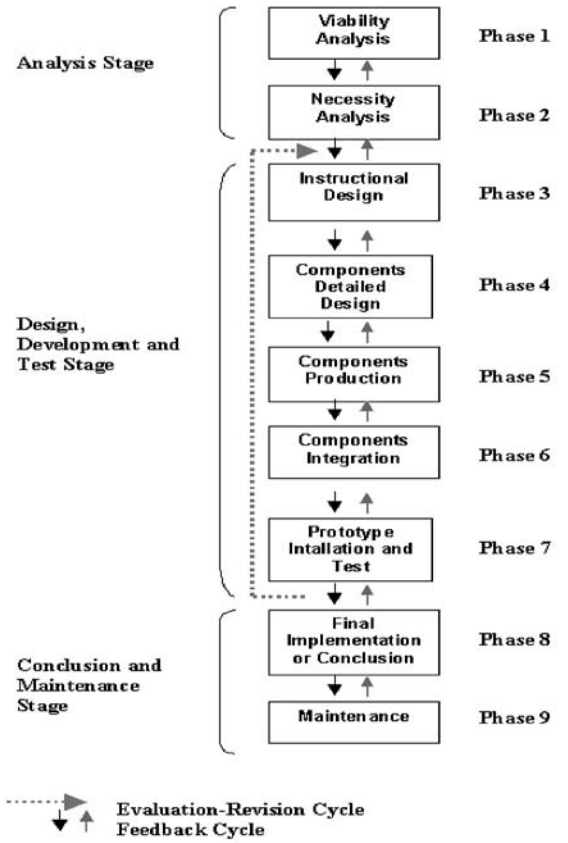
The model used for the planning, development and evaluation of the OMEA was a result of the integration of the model presented in the first author's Masters thesis (Prata, 2000; Prata et al., 2002) with methods (Sutcliffe, 1999), methodologies and guidelines proposed by other authors (Drenner, 1998; Driedger, 1999; Fernandez, 2000; Lynch et al., 1999; McGloughlin, 2001; Nielsen, 2000; Olsina et al., 1999; Salmon, 2000; Tsai, 2000; Vaughan, 1998; Vrasidas, 2000). The final model is composed of nine phases grouped in three different stages and, in very general terms, marks the initial development of a prototype, which, if proven efficient, will serve as the basis for the subsequent development of the final OMEA.

Structure of the Model

The final model, which can be seen in more detail in Figure 1, is composed of nine phases grouped in three different stages, namely: the analysis stage (phases of viability and necessity analysis), the design, development and test stage (instructional design, components detailed design, components production, components integration and prototype installation and test phases) and, finally, the conclusion and maintenance stage (final implementation or conclusion, and maintenance phases).

In general terms, the model marks the initial development of a prototype, which, if proven efficient, will serve as the basis for subsequent development of the final educational application. The software evolves according to an organized sequence of phases, in which the output of a given phase constitutes the input of the following phase. Each of these phases is composed of a varying number of tasks, which should be totally completed before moving onto the next phase. In an ideal situation the software would be developed following all phases of the model, in the exact order presented, and an efficient prototype would probably be obtained on the first trial. Unfortunately, things do not usually work out this well in practical terms. In reality, what happens is that it is frequently necessary to return to previous phases. These situations are also foreseen in the model, namely through the presence of feedback cycles in all phases. The tests or evaluations occur during

Figure 1. Detailed structure of the model



all phases and constitute the so-called formative evaluation. The final tests or summative evaluation only occur at the end of phase 7 and are meant to test the efficiency of the developed prototype. If the prototype is confirmed as efficient, then the remaining models are implemented following the same process. If the prototype is not efficient then we return to phase 3 of the model, or rather, we enter a evaluation-revision cycle.

It was noted that in all analyzed development models (lifecycle models and models for the development of educational multimedia software) there are phases and tasks, which, though bearing different designations, are common to most of the models. After analyzing each of these phases, their relevance to this work and the order in which they should be included in relation to the other phases, we reached the following final result:

1. Phase 1 – Viability Analysis
   - **Goal:** Analyze the viability of the project.
   - **Input:** Students' needs.
   - **Tasks:** Define without going into great detail:
     - The goals; the target population; the subject or domain; the choice of medium; tasks, resources, constraints and costs; alternative implementation strategies
   - **Output:** A viability report considering the alternatives.
2. Phase 2 – Necessity Analysis
   - **Goal:** Description of the situation and of the learning goals.
   - **Input:** Initial needs identified in the viability report.
   - **Tasks:** Detailed defining:
     - The make-up of the working and executive groups; a survey of field conditions; target population descriptions (knowledge, motivation, capacities, objectives and exceptions) and the choice of evaluation groups; learning necessities analysis and a hierarchy of learning objectives; subject analysis, problem description, objective and generic selection of components; selection of development environment (educational software model); analysis and multimedia tools selection.
   - **Output:** Needs specification; contents learning/educational structure (learning model).
3. Phase 3 – Instructional Design
   - **Goal:** Conceptual instructional design.
   - **Input:** Specific needs and learning model.

- **Tasks:** Instructional design (navigation, colors, background and feedback among others); contents structuring (division into hierarchical levels and prototypical schemes); instructional events identification and sequence (navigation maps, for instance).
- **Output:** General schemes of content structure and sequence of instructional events.

4. Phase 4 – Components Detailed Design
   - **Goal:** Detailed design the several structures.
   - **Input:** General schemes of content structure and sequence of instructional events.
   - **Tasks:** Creation of a detailed design of the several structures (general base layouts); also, creation of a detailed design of instructional events (base layouts for instructional events).
   - **Output:** Functional specification (detailed design of the base interface).
5. Phase 5 – Components Production
   - **Goal:** Collect all the components that are to be used.
   - **Input:** Functional specification.
   - **Tasks:** Distribute contents through base layouts; identification and detailed study of said components (general design of the prototype's interface); collect and/or produce the different components (editorial work) and script (detailed design of prototype's interface); proceed to final selection of the authoring system.
   - **Output:** All multimedia components that are to be used, alternatives, and script (detailed design of the prototype's interface).
6. Phase 6 – Components Integration
   - **Goal:** Conclude the prototype.
   - **Input:** Functional specification script and all multimedia components that are to be used.
   - **Tasks:** The integration of all collected and/or produced components.
   - **Output:** Concluded prototype.
7. Phase 7 – Prototype Installation and Test
   - **Goal:** To install the prototype, evaluate its efficiency and identify weak points.
   - **Input:** Concluded prototype.
   - **Tasks:** Prototype installation; prototype evaluation (to detect errors and evaluate its efficiency).
   - **Output:** The results of the prototype evaluation.
     - **NOTE:** If the prototype is not efficient, it is necessary to go back to the instructional design phase in order to proceed to its adjustment. This process is called the evaluation-revision cycle and it does not imply that all phases have to be repeated with the same degree of detail. Depending on the types of weak points and/or faults identified during the evaluation it is necessary to ascertain which phases need revision.
8. Phase 8 – Final Implementation or Conclusion
   - **Goal:** Conclude the educational application.
   - **Input:** Documents resulting from all previous phases, as well as the prototype.
   - **Tasks:** The correction of faults detected during the prototype's evaluation; the incorporation of any remaining modules in the prototype.
   - **Output:** Final educational application.

9. Phase 9 – Maintenance
    - Regular maintenance of the educational contents of the application is necessary in order to keep it up to date.

As a working philosophy, we decided to adopt an inherent characteristic of the Carrie Heeter model, that is, letting creativity decide the course of the software's design (Scarlatos, 1997).

The evaluation referred to in phase seven concerns the final tests to be conducted on the prototype. However, and even though this is not schematically represented in the model, intermediate tests should be carried out for all phases.

EVALUATION METHOD

Each one of the four developed prototypes underwent two kinds of evaluation:

- Formative Evaluation – which was based on the tests conducted during the entire development process (intermediate tests). This kind of evaluation was carried out by a team of two people from the work team, 10 people belonging to the target population, and four people belonging to a potential target population, that is, students from another Higher School, who also take multimedia classes. The goal of this kind of evaluation was to identify and correct problems during a preliminary phase and at the same time, to understand the students' expectations and preferences;
- Final Evaluation – which consisted of a final test or experiment and also of the handing out of a questionnaire, as will be described in the next subsections.

Final Test or Experiment

Students from the target population participated in this experiment, the goal of which was to evaluate the efficiency of the prototype as a substitute for the *Organizational Multimedia* face to face classes. The experiment is described below:

- As mentioned in previous sections, the *Organizational Multimedia* classes are laboratory-based and last three hours: one and a half hours are dedicated to theory and the other hour and a half is set aside for practice. Thus, to begin with, all students participated in a one and a half hour theory-based face to face class on the theme covered by the prototype.
- After that, there was an attempt to place the students in four categories: students with experience using the Internet (and with experience using the multimedia components used on the prototype), students with experience using the Internet (and with no experience using the multimedia components used on the prototype), students with no experience using the Internet (and with experience using the multimedia components used on the prototype) and students with no experience using the Internet (and with no experience using the multimedia components used on the prototype).
- After being placed in categories, the students were then distributed into two groups. Each group comprised the same number of students from different categories and occupied a separate classroom.
- One of the groups was submitted to the other hour and a half face to face class, which corresponded to the practice-based class (practical part of the laboratory-based class). Meanwhile, in the other group, each student for an hour and a half had access to the prototype which was meant to act as the substitute for the practice-based face to face class being attended by the others. This group of students used the prototype under the supervision of a teacher in order to guarantee obedience to the established rules.

During the practice-based face to face class we tried not to convey more essential information than was covered by the prototype. Thus, the students who used the prototype did not access less essential information than those who attended the practice-based face to face class did. What they did have, however, was access to a different teaching method with different characteristics.

- The prototype was installed on a public server and each user achieved speeds similar to those of a 56Kb-modem connection (the most frequent type of connection speed amongst the student population).
- Afterwards, all students were submitted to the same individual practical exercises.
- The next step was the analysis of results.

The students chosen were considered natural because they were in the last year of their degrees and, at the time, all of them were taking the *Organizational Multimedia* course.

By separating the student population into two groups and submitting each group to a specific situation (one of the groups having only studied using the prototype and the other group having attended the practice-based face to face class), we used an experimental methodology of the *between-groups* type, which seemed to most appropriate (Santos et al., 1999).

However, with the goal of improving the experiment, we considered the possibility of performing it in two different phases:

A first phase would be exactly what is described above and a second phase would involve changing groups: having the group that initially studied with the prototype now attend the practice-based face to face class, and having the group that initially attended the practice-based face to face class now study with the prototype. The studied contents could be:

- The same in both phases of the experiment, but in this case, it would not be the first time that the students were exposed to the subject matter, which would therefore carry a learning situation from the first phase of the experiment to the second. To solve this type of situation, what we could do would be to carry out the second part of the experiment after some time. However, that solution would have been impossible in this case as the students, who were in the last year of their courses, were about to finish school and obviously were not very committed to coming back later to participate in a new experiment. However, even if the experiment were possible, the experience of these students in using the Internet would not be the same, because the majority of them have easy access to these technologies and strong motivation to use them.
- Different in two phases of the experiment – if the contents used on the second phase of the experiment were different from the ones used in the first phase, then the second phase of the experiment could be conducted immediately after the first one. However, as the complexity level of the contents varies from theme to theme, and as the multimedia structure of the different themes is also different, the result will not be reliable. Another possibility would be to consider the population as an individual group and submit them to both situations: all students would be supposed to attend the practice-based face to face class and after that use the prototype, or vice versa. This methodology, of the "inside-groups" type, did not seem to be very appropriate because, from one situation to another, we will have situations of learning.

We can therefore conclude that the options referred to above work as a new experiment, similar to the first one, but not as a complement to it.

Questionnaire

The second part of the final evaluation was the use of a questionnaire. We were expecting to capture the students' reactions to the prototype, find out their opinions about it, detect technical failures, and collect some suggestions in order to improve it. As we wanted to include all the students in this evaluation, we had to partially repeat the final test or experiment. That is to say, after finishing the practical exercises, the groups were switched around. The group that had initially studied with the prototype attended a practice-based face to face class, and the group that had initially attended the practice-based face to face class studied with the prototype. After that, they were all able to voice an opinion about the prototype's characteristics and fill out the questionnaire.

We asked them to give their opinion freely (in writing and anonymously) concerning all details that they particularly liked and disliked. The students were encouraged to navigate around all the prototype's paths. No information was given (not even a manual or any previous tips) about using the prototype. While the students were using the prototype, they were observed in a passive manner. This gave us the opportunity to collect information about their reactions, expressions and behaviors.

RESULTS OBTAINED

Prototype with a Class on Image

In the final evaluation of the prototype, which focused on a practice-based class about image, 54 students participated (38 of whom with previous experience in using the Internet and 16 with no experience using the Internet). All the students experienced in using the Internet also had experience using image files). After the experiment and the submission of the questionnaire, the results were analyzed as follows:

Experiment Results

In relation to the experiment, in general terms, all students solved the exercises easily. However, the results obtained were different, as presented in Table 1.

In both situations (students with experience and students with no experience using the Internet) we tried to verify if the variables of grade obtained in the exercises and method of study used were or not co-related (dependent). As the sample size was considered small, we used Fisher's Test (Everitt, 1997) and we verified that the variables were not independent. That is to say that, in both situations, the grade obtained in the exercises was not independent of the study method used.

The experiment showed us that the majority of the students feel quite enthusiastic about using this type of online application and that its use may

Table 1. Average grades obtained by the students

	Students who initially used the **prototype**	Students who initially attended the **practice-based class**
With experience using the Internet and image files (**38** students)	**14**	13
With no experience using the Internet (**16** students)	13	13

have better results than previously expected. In fact, as this application is meant to be a substitute for attending classes, the main goal is to achieve the same results that we achieve with the classes, and this happened in the group of students with no experience in using the Internet. However, the results obtained in the group experienced in using the Internet were a little better amongst students who initially only used the prototype than the results obtained amongst those who initially attended the practice-based class (at an average of 14 against an average of 13). This probably happened because the class was about image, which is simple to represent and did not occupy too much space, when compared with more complex media such as sound, animation or video. In fact, the prototype, which was very easy to use and relatively fast, worked very well in terms of motivation. These results prove the efficiency of the prototype and are a good incentive to continue this work.

Questionnaire Results

As for the questionnaire, the direct answers are summarized as presented in Table 2.

It was noted, as expected, that the 38 students with experience in using the Internet were more demanding and, in fact, they are amongst the ones who classified the prototype as slow. Another factor was that although all students considered this type of prototype to be a good (51 out of 54) or medium (three out of 54) substitute for attending classes, only 45 out of 54 considered the studied prototype good enough to replace the class entirely. This probably indicates that there are still

Table 2. Answers obtained on the questionnaire

About the prototype:	Very	More Or Less	A Little	Very Little
1. Has an attractive design	42	11	1	0
2. The information is well organized	45	7	2	0
3. Navigation is simple and intuitive	37	10	7	0
4. The subject is clearly presented	49	5	0	0
5. Easy to use	44	10	0	0
6. Motivating	50	4	0	0
	Good	Medium	Weak	Poor
7. Number of examples presented	51	3	0	0
8. Number of exercises presented	41	10	3	0
9. FAQs section	53	3	8	0
10. Speed	29	16	9	0
11. Constitutes good method of replacing attended class	51	3	0	0
12. Global evaluation	40	11	3	0
	Yes		No	
13. A good enough substitute for the attended class?	45		9	

some improvements to be made to the prototype. In relation to the open-ended questions, students were asked to identify the strong points, the weak points and to suggest ways of improving the prototype. The strong points mentioned were that it helped students who could not attend classes (50 out of 54), that it was accessible from anywhere (51 out of 54), that it was motivating (49 out of 54), enjoyable (32 out of 54) and different from other study methods (43 out of 54). The weak points mentioned were average navigational structure (16 out of 54) and slowness (23 out of 54).

The majority of students considered that the prototype had an attractive design, that the information was well organized, that navigation was simple and intuitive, that subjects were clearly presented, that it was easy to use and motivating, that it presented an adequate number of examples and exercises and that it was a good substitute for attending classes. In fact, the only problems pointed out were a degree of slowness in downloading the prototype and an average navigational structure.

Prototype with a Class on Sound

In the final evaluation of the prototype, which included a practice-based class on sound, we had the participation of 82 students (16 students experienced at using Internet and sound files, 40 students experienced at using the Internet but with no experience of using sound files and 26 students with no experience of using Internet or sound files). After conducting the experiment and applying the questionnaire, the results were analyzed as follows.

Experiment Results

As for the experiment, in general terms, all students solved the exercises easily. However, the results obtained were different, as presented in Table 3.

With Fisher's test (Everitt, 1997) we verified that the variables (grade obtained in the exercises and study method used) were indeed co-related (not independent). That is to say that the grade obtained in the exercises did relate to the study method used.

As for the results obtained by students who initially attended the practical class and results obtained by students who initially used the prototype, there were some differences, namely:

- The results obtained among students with no previous experience of using the Internet

Table 3. Average grades obtained by the students

	Students who initially used the **prototype**	Students who initially attended the **practice-based class**
With experience using the Internet and sound files (**16** students)	14	**16**
With experience using the Internet but with no experience using sound files (**40** students)	13	13
With no experience using Internet and with no experience using sound files (**26** students)	12	12

or sound files (26 out of 82) and the results obtained among students with experience of using the Internet but with no experience of using sound files (40 out of 82) were more or less the same for both groups.

- The results obtained among students experienced in using the Internet and sound files (16 out of 82) were very good and even better among those who attended the class (an average of 16 against an average of 14). These students already had previous knowledge of sound subject matter and, as it was probably an interesting subject for them participated quite actively in class by asking several questions and expressing doubts, which were immediately clarified. Those who used the prototype may also have had their doubts but as they were not immediately clarified (they had to resort to the FAQ Section), their results were slightly worse.

These results were encouraging because in both groups (the group that attended the practice-based class and the group that only used the prototype) the majority of students achieved similar scores. The only category of students to achieve better averages in the group that attended the class (as compared to the average obtained by the group that only used the prototype) was the one with previous experience using the Internet and sound files (16 out of 82). However, these higher averages were the result of the personality and attitude of these particular students, who prefer asking the professor direct questions and obtaining rapid answers, rather than browsing

Table 4. Answers obtained on the questionnaire.

About the prototype:	Very	More Or Less	A Little	Very Little
1. Has an attractive design	58	18	6	0
2. The information is well organized	63	16	3	0
3. Navigation is simple and intuitive	68	12	2	0
4. The subject is clearly presented	69	13	0	0
5. Easy to use	67	13	2	0
6. Motivating	79	3	0	0
	Good	Medium	Weak	Poor
7. Number of examples presented	75	7	0	0
8. Number of exercises presented	50	22	10	0
9. FAQs section	72	5	5	0
10. Speed	4	34	36	8
11. Constitutes good method of replacing attended class	78	4	0	0
12. Global evaluation	51	30	1	0
	Yes		No	
13. A good enough substitute for the attended class?	68		14	

around the prototype looking for the answer. We consider that this problem can be solved with a more friendly and personalized FAQ Section.

Questionnaire Results

The answers are summarized in Table 4.

As we can see from Table 4, some attention needs to be paid to the results relating to the prototype lack of speed. It was noted, as expected, that the 56 students experienced in using the Internet were more demanding and, in fact, were among those who classified the prototype as slow. Another discovery was that although all students considered this type of prototype to be a "good" (78 out of 82) or "average" (four out of 82) substitute for attended classes, only 68 out of 82 considered the studied prototype good enough to replace the class. This indicates that there are still some improvements to be made to the prototype in order to make it more efficient and comparable to the attended classes.

In relation to the open-ended questions, students were asked to identify the strong points and weak points of the prototype and to suggest ways of improving it. The strong points mentioned were: the fact that it helps students who cannot attend classes (80 out of 82); it is accessible from anywhere (75 out of 82); it is motivating (79 out of 82) and it is a novel and original study method (73 out of 82). The only weak point mentioned was slowness (44 out of 82).

We saw that the students believed the prototype gave them autonomy (80 out of 82), flexibility (75 out of 82) in regard to their studies, and classified it as a motivating (79 out of 82) and novel (73 out of 82) teaching/learning process.

Prototype with a Class on Animation

In the final evaluation of the prototype, which included a practice-based class on animation, 68 students participated (16 students with Internet and animation file experience, 34 students with Internet experience but with no experience in using animation files and 18 students with no experience of using the Internet or animation files). After conducting the experiment and applying the questionnaire, the results were analyzed as follows:

Experiment Results

As for the experiment, in general terms, all students solved the exercises easily. However, the results obtained were different, as presented in Table 5.

With Fisher's test (Everitt, 1997), we saw that the variables (grade obtained on the exercises and study method used) were indeed co-related

Table 5. Average grades obtained by the students.

	Students who initially used the **prototype**	Students who initially attended the **practice-based class**
With Internet and animation file experience (**16** students)	14	**15**
With Internet experience but with no experience using animation files (**34** students)	13	13
With no experience of the Internet or animation files (**18** students)	11	11

(dependent). That is to say that the grade obtained on the exercises did relate to the study method used.

As for the results obtained by students who initially attended the practice-based class and results obtained by students who initially only used the prototype, there were some differences, namely:

- The results obtained by students with no Internet experience (18 out of 68) were the same in both groups.
- The results obtained by students with Internet experience but with no previous knowledge of animation (34 out of 68) were more or less the same for both groups.
- The results obtained by students with Internet experience and previous knowledge of animation (16 out of 68) were very good, and better among those who sat through the attended class. These students already had previous knowledge of animation and participated quite actively in class by asking questions and expressing doubts, which were immediately clarified. Those who used the prototype may also have had questions but as these were not immediately clarified (they had to use the FAQs section) the result was a slightly worse score.

These results were optimistic because in both groups (the one that attended the practical class and the one that only used the prototype) the majority of students achieved similar scores. The only category of students to achieve a better average in the group that attended the class, when

Table 6. Answers obtained on the questionnaire

About the prototype:	Very	More Or Less	A Little	Very Little
1. Has an attractive design	47	18	3	0
2. The information is well organized	60	7	1	0
3. Navigation is simple and intuitive	58	8	2	0
4. The subject is clearly presented	61	7	0	0
5. Easy to use	56	10	0	0
6. Motivating	63	5	0	0
	Good	Medium	Weak	Poor
7. Number of examples presented	60	7	1	0
8. Number of exercises presented	19	36	13	0
9. FAQs section	57	6	5	0
10. Speed	0	21	37	10
11. Constitutes good method of replacing attended class	59	9	0	0
12. Global evaluation	44	15	9	0
	Yes		No	
13. A good enough substitute for the attended class?	49		19	

compared with the average obtained by the group that only used the prototype, was the one with previous Internet and animation file experience (16 out of 68). As with the prototype with the class on sound, we believe that this higher average was the result of the personality and attitude of these particular students, who prefer asking the professor direct questions and thus obtaining rapid answers, rather than browsing around the prototype looking for answers. Probably, this problem will be solved with a more friendly and personalized FAQ Section.

Questionnaire Results

The answers are summarized in Table 6.

As we can see from Table 6, the majority of students considered the prototype "good" in general terms. However, some attention needs to be paid to the results relating to the prototype lack of speed. As with the previous prototypes, we also noted that although all students considered this type of prototype to be a "good" (59 out of 68) or "average" (nine out of 68) substitute for attending classes, only 49 out of 68 considered it good enough to replace the class. This indicates that there are still some improvements to be made to the prototype in order to make it more efficient and comparable to attending class.

In relation to the open-ended questions, students were asked to identify the strong points and the weak points of the prototype and to suggest ways of improving it. The strong points mentioned were the fact that it helped students who could not attend classes (58 out of 68), that access was possible from anywhere at any time (61 out of 68), that it was motivating (62 out of 68) and that it was a novel and original study method (36). The only weak point mentioned was slowness (47out of 68).

Prototype with a Class on Video

In the final evaluation of the prototype, which involved a practice-based class on video, 32 students participated (14 students with Internet experience and 18 students with no Internet experience). None of them had any previous experience of video files. As the size of the sample population was very small (because we were at the end of the semester), the results should be considered no more than a mere indicator, useful essentially for improving the prototype.

After conducting the experiment and applying the questionnaire, the results were analyzed as follows:

Experiment Results

As for the experiment**,** in general terms, all students solved the exercises easily. However, the results obtained were different, as presented in Table 7.

With Fisher's test (Everitt, 1997) we saw that the variables (grade obtained in the exercises and study method used) were indeed co-related

Table 7. Average grades obtained by the students

	Students who initially used the **prototype**	Students who initially attended the **practice-based class**
With Internet experience (**14** students)	12	**13**
With no Internet experience (**18** students)	10	**12**

(dependent). That is to say that the grade obtained on the exercises did relate to the study method used.

In general terms, both groups solved the exercises easily. However, in both groups, the results obtained among students who initially only used the prototype were worse than the results obtained among those who initially attended the practice-based class. These inferior results were especially noticeable among the students with no Internet experience. It was noted, as expected, that the 14 students with Internet experience were more demanding. However, they were much more tolerant of slowness while downloading the prototype components than the students with no Internet experience.

Questionnaire Results

The answers are summarized in Table 8.

As we can see from Table 8, the majority of students considered that the prototype had an attractive design, that information was well organized, navigation was simple and intuitive, that subjects were clearly presented, that it was motivating, that it presented a sufficient number of examples and exercises, and that it was a good substitute for attending classes. However, although all students considered this kind of prototype to be a "good" (27 out of 32) or "average" (five out of 32) substitute for attending classes, only 13 out of 32 considered the studied prototype good enough to substitute for the class. This indicates that there

Table 8. Answers obtained on the questionnaire

About the prototype:	Very	More Or Less	A Little	Very Little
1. Has an attractive design	25	4	3	0
2. The information is well organized	16	10	6	0
3. Navigation is simple and intuitive	11	17	4	0
4. The subject is clearly presented	23	9	0	0
5. Easy to use	3	11	18	0
6. Motivating	21	10	1	0
	Good	Medium	Weak	Poor
7. Number of examples presented	28	4	0	0
8. Number of exercises presented	26	6	0	0
9. FAQs section	24	6	2	0
10. Speed	0	9	17	6
11. Constitutes good method of replacing attended class	27	5	0	0
12. Global evaluation	3	16	13	0
	Yes		No	
13. A good enough substitute for the attended class?	13		19	

are still serious improvements to be made to this prototype in order to make it more efficient.

As for the open-ended questions, students were asked to identify strong and weak points and to suggest ways of improving the prototype. Strong points mentioned were the fact that it helped students who could not attend classes (20 out of 32), that it could be accessed from anywhere (26 out of 32), that it was motivating (17 out of 32) and that it was a novel study method (27 out of 32). The weak points mentioned were difficulties in using the prototype (25 out of 32), average navigational structure (15 out of 32) but, especially, slowness (24 out of 32).

From observing the students directly, we were able to see a great deal of enthusiasm when they started using the prototype. However, after some time they lost part of their enthusiasm and showed some impatience with the prototype's slowness.

For the first time in the course of these experiments, the results obtained among students (of all categories) who initially used the prototype were worse than the results obtained among those who initially attended the practice-based class. This probably happened because the class was about video, which is very difficult to represent/implement and resulted in a slow prototype. In this particular case, with all categories of students, the face to face class was more effective than the application.

FUTURE TRENDS

Regarding future work, we expect to conclude the implementation of the final OMEA with all the classes of the *Organizational Multimedia* course. This final OMEA, which is currently being implemented, as mentioned before is meant to substitute, in the best possible way, the face to face classes students could not attend. During the first phase the OMEA will be made available on the Internet, meaning it can be accessed by anyone, anywhere and at anytime (as an example, please check MIT's open courses, available at *http://ocw.mit.edu/index.html*). As the OMEA's main goal is to be an efficient substitute for the face to face classes, its contents are very complete and detailed, as to exactly reproduce the subjects delivered in those classes. The only thing that is different is the way in which the information is presented, because, obviously, it is adapted to be displayed in an online educational application. With this application, the student will be able to access from anyplace at anytime the content corresponding to classes that he or she missed. Creating this OMEA was the quickest and most efficient solution found to solve problems/difficulties inherent to student absences.

The Higher School of Management Sciences currently has several projects in place, namely, the implementation of an intranet and an e-learning system. Thus, and considering the OMEA's future integration in both systems, its contents are being developed in a modular fashion. This kind of development will allow the OMEA's integration with both projects' platforms and will also permit different ways to associate the contents/modules considering the required level of difficulty in each case.

Concluding, the goal is that in the future the OMEA is to be integrated and made available, in the following order, through the following systems:

- The Internet (which will help us to provide an educational environment that we classified as "online anytime, anyplace, anybody" model).
- The school's intranet (which will help us to provide an educational environment that we classified as "online in the studyplace" model).
- The school's e-learning system, and in that case it will have to be adapted to the teaching model being used (blended or exclusively online).

- Other standardized e-learning systems (LMS). In order to achieve this goal, the modules will have to be portable (multi-platform), interoperable and reusable, which is, according to Elliot Masie from the Masie Center, the present market tendency (Masie, 2004). However, and as this standardization process is already in a early stage and far from achieve universal and wide-world accepted standards (Holley, 2003), the integration of our application with other standardized systems will be the last part of the entire work.

The following describes the way in which the OMEA is supposed to be used:

1. **As a complete online substitute for the face to face classes students could not attend.** Pedagogically speaking, this is a learning model where students are at the center of the model, and we can call it online "anytime, anyplace, anybody". Some functionalities are currently being studied as to be implemented on a later stage of this work. Those functionalities include: "virtual post-its" that allow students to add their own comments and notes on the contents, exercise sections including exercises with answers and others only with the final result (solution), evaluation tests section, which will be automatically corrected by the system's application, FAQs section and personalized classes. These personalized classes will be manually or automatically generated using modules in which the student has shown most weakness with. Manually – through information obtained directly from students via an online form, where he or she notes the modules found to be the hardest; or automatically – generated by the application that analyzes results obtained on exercises and evaluation tests sections. Obviously, this last solution will only be available after obtaining some input on the student through the resolution of exercises and/or evaluation tests.
2. **As a tool to help/support the face to face classes.** In this learning model, which we can call "online in a classroom," the professor will be in control and is the center of the model. In fact this is like a traditional face to face class where the online educational application is made available not only as a background material but is integrated with classroom instruction as a classroom tool.
   Apart from the motivation associated with this learning model, one of the biggest advantages is that students learn with the professor how to use the application and thus need not do it on their own.
   The application will be used in the classroom to support the following tasks:
   - Examples demonstration – through the use of the application the students will, more rapidly, easily and efficiently understand the theoretical concepts. In our OMEA, and considering that the course classes are on multimedia, with a simple click students will, for instance, hear, see and compare sound and video files recorded using different quality parameters; see and compare the visual differences obtained after resizing images and after resizing graphics; compare the quality of an image recorded with different color palettes and compression algorithms; and so on.
   - Exercises resolution – through the use of the exercises section.
   - Evaluation tests – through the use of the evaluation tests section.
3. **In students' evaluation processes.** Through the evaluation tests section, and similarly to what can be achieved with the WebCT, once all students are connected, the application will allow them to solve the same test.

However, and in order to avoid any cheating, the questions will be presented in a different order to every one of them. The tests will be automatically corrected by the system's application, which brings some obvious advantages: the results will be quickly known and the professor will have more time to spend with more important tasks such as supporting and helping students with their learning process.

4. **As a regular study tool/support.** Which will be very useful for students because the application includes an exercise section, evaluation tests section and a FAQ's section.

   Another advantage may be what we have decided to call the "dejá vu learning advantage": the application's contents are made available in the exact same order and level of complexity as the contents presented at the face to face classes, thereby making it easier to remember and memorize these when studying.

CONCLUSIONS

Mainly in order to make information easier for students to access at any time from any place, we decided to develop an Online Multimedia Educational Application (OMEA) to aid in the teaching of the *Organizational Multimedia* course. Others goals were to benefit from the advantages of using information technologies (namely Web-based technologies) as a teaching support inside and outside the classroom, to contribute to the modernization of study methods used by students and, finally, to acquire more experience in planning, developing and evaluating OMEA. The *Organizational Multimedia* course is taught to students in their fourth year at the Higher School of Management Sciences, where the first author teaches. In order to plan, develop and evaluate this type of application, the use of adequate models is highly recommended. The model followed and described in this chapter involves the initial development of a prototype, which, if proven efficient, will serve as the basis for subsequent development of the final application. As the *Organizational Multimedia* course is made up of six different chapters: text, image, graphics, sound, animation and video, our first thought was to develop a prototype for each chapter. However, after more detailed analysis, we noticed that some chapters include components with similar characteristics. Thus, the final decision was to develop four prototypes in the following order: a first prototype with a practice-based class on image, a second one with a practice-based class on sound, a third one with a practice-based class on animation, and finally, a fourth one with a practice-based class on video.

As part of a major project, the model used to plan, develop and evaluate the prototypes as well as the evaluation method used were the same in all four prototypes. Both the model and the evaluation method used have been described in this chapter.

The final evaluation of the prototypes was made up of two parts: an experiment and a questionnaire, both requiring the participation of different categories of students. However, as the sample size was considered medium or small in all four experiments, the results achieved should be considered merely indicative, and useful for improving the prototypes.

In general terms, the experiments (four: one for each prototype) showed us that the majority of students felt quite enthusiastic about using this type of application and that it may have better results than expected. In fact, as this application is meant to be a substitute for attending classes, the main goal is to achieve the same results we achieve with face to face classes, and that happened with the majority of the prototypes and categories of students, as we can see in Table 9.

As we can see from Table 9, in general terms, the prototypes were efficient with 70% (54.5% +

Table 9. Results obtained for all four prototypes.

	Number of students whose results with the prototype were worse than results achieved with the face to face class	Number of students whose results with the prototype and the face to face class were the same	Number of students whose results with the prototype were better than results achieved in face to face class
Prototype with a class on image: (tested by 54 students)	0 (0%)	**16 (29.6%)**	**38 (70.4%)**
Prototype with a class on sound: (tested by 82 students)	16 (19,5%)	**66 (80.5%)**	0 (0%)
Prototype with a class on animation: (tested by 68 students)	16 (23,5%)	**52 (76.5%)**	0 (0%)
Prototype with a class on video: (tested by 32 students)	**32 (100%)**	0 (0%)	0 (0%)
Total (246):	**64 (26%)**	**134 (54.5%)**	**38 (15.5%)**

15.5%) of the students. However, when we consider each prototype separately, we can see that:

- The prototype with a class on image worked even better than expected. In fact, 70.4% of the students achieved better results with the prototype than by attending face to face class.
- The prototype with a class on sound needs to be improved. There is a particular category of students (those with Internet and sound file experience – 19.5%) who achieved better results by attending the face to face class than with the prototype. As this category of students is more demanding and dependent upon direct student-teacher interaction, one solution may be the improvement of the FAQ section.
- The prototype with a class on animation needs to be improved. There is a particular category of students (the ones with Internet and animation file experience – 23.5%) who achieved better results by attending the face to face class than with the prototype. As this category of students is more demanding and dependent upon direct student-teacher interaction, one solution may be the improvement of the FAQ section.
- The prototype for a class on video needs to be completely reformulated because 100% of the students obtained worse results with the prototype than by attending the face to face class.

It was possible to note from the questionnaires that the majority of students considered the prototypes to have an attractive design, to contain well organized information, to feature simple and intuitive navigation, to present subjects clearly, to be easy and motivating, to contain a sufficient number of examples and exercises, and to be a good substitute for attending classes. As for weak points, it was also possible to obtain enough feedback to improve the prototypes. In fact, the only problems consistently reported were slowness (in all the prototypes) and only average navigational structure in two prototypes (the prototype with a class on image and the prototype with a class on

video). In the prototypes with classes on sound and animation we felt that the FAQs sections had to be improved in order to become more user-friendly and personalized for a particular category of more demanding students, and those more dependent on traditional student-teacher interaction.

The results obtained were very encouraging and showed us that the production of this type of online applications should be encouraged. The enthusiasm and the results achieved by using the prototypes justify further development work, and at present the prototypes that were already perfected are being now used in the implementation of the final OMEA. As the Higher School of Management Sciences has currently several projects in place, namely, the implementation of an intranet and an e-learning system, the final OMEA contents are being developed in a modular fashion in order to be easily integrated with both systems. So, the goal is that in the future, the final OMEA may be integrated and made available, in the following order, through the following systems: the Internet, the school's intranet, the school's e-learning system and, when possible, in other standardized e-learning systems (LMS). As to the way in which the OMEA is going to be used: as a complete online substitute for the face to face classes students could not attend, as a tool to help/support the face to face classes (in the following tasks: examples demonstration, exercises resolution and evaluation tests), in students' evaluation processes and as a regular study tool/support.

Concluding, in a general way the use of these types of OMEA, inside or outside the classroom, brings considerable benefits to its users (students and professors), namely:

- Students who cannot attend the classes have access to the exact same contents presented at those classes without depending on others for help.
- Contents are available to anybody at any time from any place.
- Students easily understand theoretical concepts when associated with practical demonstrations via multimedia files with images, sounds, animation and video.
- Students in face to face classes do not need to be constantly concerned on taking notes all the time and may therefore instead pay more attention, dedication and immerse themselves in the learning process itself, by analyzing practical cases, solving exercises or simply discussing the subjects with the professor and colleagues.
- Professors have more time and feel more free to attend to students' solicitations and doubts;
- Students feel more motivated not only because they are using information technologies inside and/or outside the classroom but also because they are using real online multimedia educational applications and because their study methods are being modernized.

In conclusion, this was a worthwhile project since the goals were achieved. Considering that the development of this type of application is a complex process, the results obtained with the prototypes of the OMEA were very good, indeed better than expected. This project also allowed us to see that students are enthusiastic about learning through this type of applications and are prepared to use them.

A fundamental contribution to the prototype's success was the use of the educational software model described in this document, which for its simplicity of use and intuitive design we consider important to recommend.

Finally, with this work we contributed to the modernization of the study methods used by students and we have acquired more experience in planning, developing and evaluating Online Multimedia Educational Applications.

REFERENCES

Abbey, B. (2000). *Instructional and cognitive impacts of Web-based education*. Hershey, PA: Idea Group Publishing.

Aggarwal, A. (2000). *Web-based learning and teaching technologies: Opportunities and challenges*. Hershey, PA: Idea Group Publishing.

Azevedo, B. (1997). Tópicos em construção de software educacional. *Estudo Dirigido.*

Berk, R., & Kanfer, A. (1996). *Review of Web style guides.* NCSA – National Center for Supercomputing Applications – Technology Research Group, University of Illinois, USA. Retrieved January 27, 2004, from *http://archive.ncsa.uiuc.edu/edu/trg/styleguide/*

Chambel, T., Bidarra, J., & Guimarães, N. (1998). Multimedia artefacts that help us learn: Perspectives of the UNIBASE Project on distance learning. *Proceedings of the Workshop on Multimedia and Educational Practice, ACM Multimedia'98*, Bristol, UK.

Chute, A., Thompson, M., & Hancock, B. (1999). *The McGraw-Hill handbook of distance learning.* New York: McGraw-Hill.

DISA. (1995). *Multimedia technology standards assessment version 2*. Prepared for the Defense Information System Agency. Retrieved May 25, 1998, from *http://www.ott.navy.mil/refs/stds/mtsa/*

Drenner, D. (1998). *Audio, video and digitizing sound and video clips for various language courses.* Toronto: Multimedia Lab of University of Toronto, CHASS, Toronto, Canada. Retrieved February 2, 1999, from *http://lab.chass.utoronto.ca/Damion*

Driedger, J. (1999). *Multimedia instructional design*. University of Alberta Faculty of Extension. Academic Technologies for Learning, Canada. Retrieved February 2, 1999, from *http://www.atl.ualberta.ca*

Everitt, B. (1997). *The analysis of contingency tables.* Chapmann & Hall.

Fernandez, J. (2000). Learner autonomy and ICT: A Web-based course of English for Psychology. *Educational Media International, 37,* 257-261.

Guimarães, N., Chambel, T., & Bidarra, J. (2000, October). From cognitive maps to hypervideo: Supporting flexible and rich learner-centered environments. *IMEJ-Interactive Multimedia Electronic Journal of Computer-Enhanced Learning.*Retrieved October 10, 2003, from *http://imej.wfu.edu/*

Hartley, K. (1999). Media overload in instructional Web pages and the impact on learning. *Educational Media International, 36*, 145-150.

Hooley, A. (2003). *Standards in e-learning.* EPIC white paper. Retrieved January 26, 2004, from *http://www.epic.co.uk/*

Könyves-Tóth, E., Megyesi, L., & Molnár, I. (1995). Methodological and psychological analysis of a multimedia educational program. *Educational Media International, 32*(1), 12-17.

LeBaron, J. (2001, August). *Online-learning in schools and higher education: An overview of thought and action, by the OFL-2001 Instructional Team.* University of Massachusetts Lowell (USA) & University of Oulu (Finland).

Lynch, P., & Horton, S. (1999). *Web style guide - Basic design principles for creating Web sites.* Yale University Center for Advanced Instructional Media.

Machado, S. (2001). *E-learning em Portugal*. Lisbon: FCA.

Masie, E. (2004). *Making sense of learning standards and specifications.* Masie Center.Retrieved January 26, 2004, from *http://www.masie.com/masie/default.cfm?page=standards*

McCarthy, P. (1995). *CAL-Changing the face of education?* CAL Research Poster, MSc Information Systems.

McGloughlin, S. (2001). *Multimedia - Concepts and practice*. New Jersey: Prentice Hall.

Nielsen, J. (2000). *Designing Web usability*. USA: New Riders Publishing.

Olsina, L., Godoy, D., & Lafuente, G. (1999). Assessing the quality of academic Websites: A case study. *The New Review of Hypermedia and Multimedia, 5,* 81-103.

Palloff, R., & Pratt, K. (2001). *Lessons from the cyberspace classroom - The realities of online teaching*. California: Jossey-Bass Inc.

Prata, A. (2000, May 2). *Planeamento e desenvolvimento de um CD-ROM para apoio ao estudo da multimédia*. Master Thesis, presented at ISCTE: Lisbon.

Prata, A. (2003). Web-based distance learning application for teaching multimedia in higher school. *Proceedings of the 4th International Conference on Information Communication Technologies in Education*, ICICTE - 2003, Samos, Greece, pp. 317-323.

Prata, A., & Lopes, P. (2002). How to plan, develop and evaluate multimedia applications – A simple model. *Proceedings VIPromCom-2002 (International Symposium on Video/Image Processing and Multimedia Communications)*, Croatian Society Electronics in Marine - Elmar. Zadar, Croatia, pp. 111-115.

Prata, A., & Lopes, P. (2003a). E-learning tool for teaching multimedia and digital video in higher education. *Proceedings of the E-Learn World Conference on E-Learning in Corporate, Government, Healthcare & Higher Education - AACE Conference*, to be held in Phoenix, Arizona.

Prata, A., & Lopes, P. (2003b). E-learning tool for teaching organisational multimedia. *e-Society 2003 IADIS International Conference Proceedings Book*, (vol. II, pp. 961-964), Lisbon, Portugal.

Prata, A., & Lopes, P. (2003c). Impact of a on-line application for teaching video to a multimedia course in higher education. *Proceedings of the ITRE 2003 - IEEE International Conference on Information Technology: Research and Education*, Newark, New Jersey, pp. 620-624.

Prata, A., & Lopes, P. (2003d). Web-based educational multimedia application for the teaching of multimedia contents: An experience with higher education students. *Information Technology and Organizations: Trends, Issues, Challenges and Solutions* (vol. II, pp. 975-976), Philadelphia, Pennsylvania.

Rosenberg, M. (2001). *E-learning - Strategies for delivering knowledge in the digital age*. New York: McGraw-Hill.

Ryan, S., Scott, B., Freeman, H., & Patel, D. (2000). *The virtual university - The Internet and resource-based learning*. London: Kogan Page.

Salmon, G. (2000). *E-moderating - The key to teaching and learning online*. London: Kogan Page.

Santos, A. (2000). *Ensino a distância & tecnologias de informação*. Lisbon: FCA.

Santos, B., & Ferreira, C. (1999). Eficácia da utilização de documentos multimedia no apoio ao estudo: Uma experiência simples. *Proceedings of CGME'99* (1ª Workshop Computação Gráfica Multimédia e Ensino).

Scarlatos, L. (1997, November). Designing interactive multimedia. *Fifth ACM International Multimedia Conference*, Brooklyn College.

Sutcliffe, A. (1999). A design method for effective information delivery in multimedia presentations. *The New Review of Hypermedia and Multimedia, 5,* 29-57.

Tsai, C. (2000). A typology of the use of educational media, with implications for Internet-based instruction. *Educational Media International, 37,* 157-160.

Vaughan, T. (1998). *Multimedia - Making it work*. California: McGrawHill.

Vrasidas, C. (2000). Principles of pedagogy and evaluation for Web-based learning. *Educational Media International, 37, 105-111.*

This work was previously published in Instructional Technologies: Cognitive Aspects of Online Programs, edited by P. Darbyshire, pp. 31-72, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 4.10
Interactive Systems for Multimedia Opera

Michael Oliva
Royal College of Music , UK

ABSTRACT

This chapter considers the development of systems to deliver multimedia content for new opera. After a short overview of the history of multimedia in opera, the specific requirements of opera are analysed, with emphasis of the fundamental musicality of operatic performance. Having considered the place of multimedia elements in the narrative and acting space, the relevance of previous practice in electroacoustic music and Vjing is considered as a model for a working approach. Several software and hardware configurations explored, including the use of gestural control by the actors themselves. The creation of a keyboard based "video instrument" with a dedicated performer, capable of integration into the pre-existing musical ensemble, is recommended as the most effective and practical solution.

INTRODUCTION

By definition, opera and musical theatre should be the ultimate multimedia experience, and new technologies are bound to become an ever increasingly important part of this art form. Indeed, I would go further and suggest that the inclusion of computer sound and video is essential to the development of the form in our new century, expanding its range hugely and bringing much needed new audiences. In particular, we need to recognize the highly visually sophisticated nature of modern audiences, who have learnt the languages of cinema and television from birth, using these techniques to enhance our storytelling in opera and creating works of contemporary relevance. It is "an honest expression of the life we're living now" as Steve Reich, an important pioneer in this field, says (Reich & Korot, 2001).

Developing suitable interactive systems will be a part of bringing new media into the fold, so that they work successfully with what is there already. At its best, the experience of live opera, both large- and small-scale work, can be overwhelmingly powerful. I am convinced that much of this power derives from the fact that it is live, and that we are aware of the fragility and variability of what we are seeing. Also, the original conception of opera as "drama through music"

still applies, not that music is the most important element, all should carry equal weighting, but it is the music that is at the helm, so to speak, and musicality has to be a key principle. It is vitally important not to lose this in the push to include computer sound and video, and this is all too easily done. In particular, the marriage of prerecorded material with live elements needs to be very carefully handled so that the performers do not feel straight jacketed, as this is immediately deadening. As far as possible, the performers and conductor must have the freedom to drive the work along and interpret it as they wish, and any system we use for delivery of the media should enable this, responding to the energy of a particular performance, and functioning in an essentially musical way.

In order to achieve a successful integration, it is just as important that the use of computer sound and video is written into the work from the outset, as an integral part of the story to be told. To use these elements purely atmospherically or decoratively is to miss the point. These additions are capable of so much more than this, creating entire layers of narrative and character, defining worlds for these characters to inhabit, and alternative timelines for them to move in. In parallel with recent instrumental electroacoustic music, the design and implementation of the system itself is a significant part of the score, with the nature of the interface between person and machine informing what is possible. Clearly, what is said will adapt itself to the nature of the medium, and this will have an effect on the content of such works and the way that they are staged.

It also has to be recognised that even small-scale opera involves the collaboration of a rather large number of people, all of them trained specialists, from conductor to stage manager, working together in, hopefully, a highly organised way. Over the centuries, chains of command have developed that are logical and effective. Therefore, careful thought must go into developing systems that are appropriate to this type of collaborative ensemble. What may work for one form of multimedia production may not work here.

Thanks to advances in computer technology, the creation of such works has become a real possibility and, drawing from my own experience of composing and producing multimedia opera and electroacoustic music, I will set out and evaluate the types of interactive systems that might be best suited to the task, exploring in detail the requirements for what is to be achieved and the software and hardware possibilities. I shall investigate the nature of a practical interface with singers, ensemble, and conductor (as well as other elements of stagecraft such as lighting and set design) and how this is to be achieved.

BACKGROUND

The use of the moving image has a venerable history in opera. During the second phase of composition of *Lulu* (1929-1934), Alban Berg included a "film music interlude" between Act II i and II ii, and inserted a "film music scenario" in the short score of the opera. The film music interlude was to accompany a silent film that continued the action, showing events that could not be shown onstage, with annotations in the score indicating how images and actions would correspond to the music, and this was realised after Bergs death at the 1937 premiere (Weiss, 1966). More recently, Tod Machover' *Valis* (1987) (Richmond, 1989), Steve Reich's *The Cave* (1993) and *Three Tales* (2002)—described by Reich as "music theatre video" works (Reich & Korot, 2001)—and Barry Truax's *Powers of Two* (1995) (Truax, 1996) represent much more complete and successful integrations of video and computer sound into operatic storytelling allowed by considerably more advanced technology.

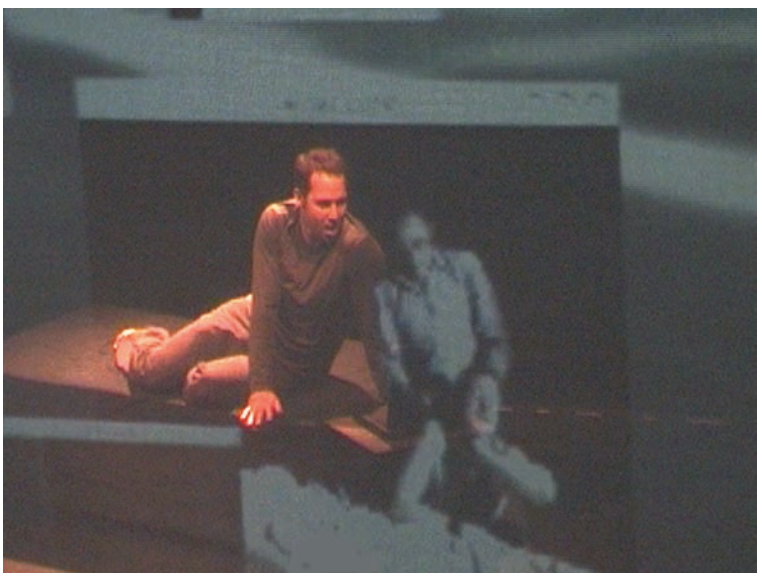
But in all these cases, video is not generated live, so the incorporation of these elements into theatrical work presents performers with a significant problem: Who owns time? Prerecorded com-

puter sound and video naturally have a tendency to "own time," running, as they are wont to do, at fixed rates. If there is to be close synchronisation, once a sequence has been triggered everyone has to play along. Moment to moment flexibility in tempo and interpretation are largely ruled out, and although there is the possibility of rubato at extremely small scales, depending on the rhythmic quality of the passage in question, there is no interactivity between events onstage and this sort of video or computer generated material.

In my own *Black and Blue* (2004) (Oliva, 2004), this problem was not resolved. The piece uses many long cued video and audio sequences and requires highly accurate synchronisation between image, action, and sound. Initially, the video represents the central character's real-time experiences viewing Web cams and in online video chatrooms, although as the drama unfolds, the narratives of the online, imagined, and real worlds become increasingly confused. The action is fairly evenly spread between these worlds, and integration was achieved by projecting the video onto a gauze screen that filled the front of the stage so that by carefully controlling lighting levels, the singers could be made to appear and disappear within the images. A second gauze at the back of the stage allowed for more complex layering. Video remained entirely in the digital domain, taking the form of a series of Quicktime .mov files on a laptop, organised into a playlist. The laptop was connected directly to a data projector, much as one would for a Powerpoint presentation. In contrast with previous tape-based systems, this allows for frame-accurate synchronisation between the image and music.

Although sequences were cued in the score to ensure sync with the live performers, the conductor worked to a click track generated by the computer, much as they might in the recording of music for film or TV. This system was simple, using commercial sequencing software that provided instantaneous response for accurate cueing. This worked, was reliable, and given the very driven nature of the story (a sort of high speed Faustian descent into hell) and score, it could be said that the overall effect was successful. However, it also proved incredibly burdensome on the performers, particularly over the large timescales involved (90 minutes).

Figure 1. A character watches an online murder in Black and Blue. The video is projected onto a gauze in front of the acting space



It could be argued that this was simply a skill for them to learn, one increasingly demanded in modern music making, and this view has some validity, but it is also the case that this method robbed talented performers of some of their most highly developed and cherished skills. To give of their best, singer/actors need to feel ownership of the music and dramatic pacing. This ownership allows them to project personality, which in turn greatly assists in the development of character. Another part of this ownership is that they need to feel safe. Mistakes can and do happen, so there should be the flexibility to allow for recovery in these situations, as is the case in more traditional acoustic performance. This safety relieves stress and lets the singer focus on the fine detail of their acting, leading to much more successful results.

I believe that a truly satisfying operatic or theatrical experience depends fundamentally on the ability of the singer/actors to shape the flow of time, using fluctuation in tempo as an expressive tool both at the level of a single phrase or gesture or over the course of an entire scene or act. My own experience in less technological theatre and music making shows that a work may show large variations in timing from night to night without it being noticeable that one performance was particularly slower or faster than another. An audience's sense of pace is a relative thing, and context is everything, so the variable, in-the-moment responses of a good performer are a key part of what we seek as members of that audience. This, above all, will draw us into the work and make us care about the characters.

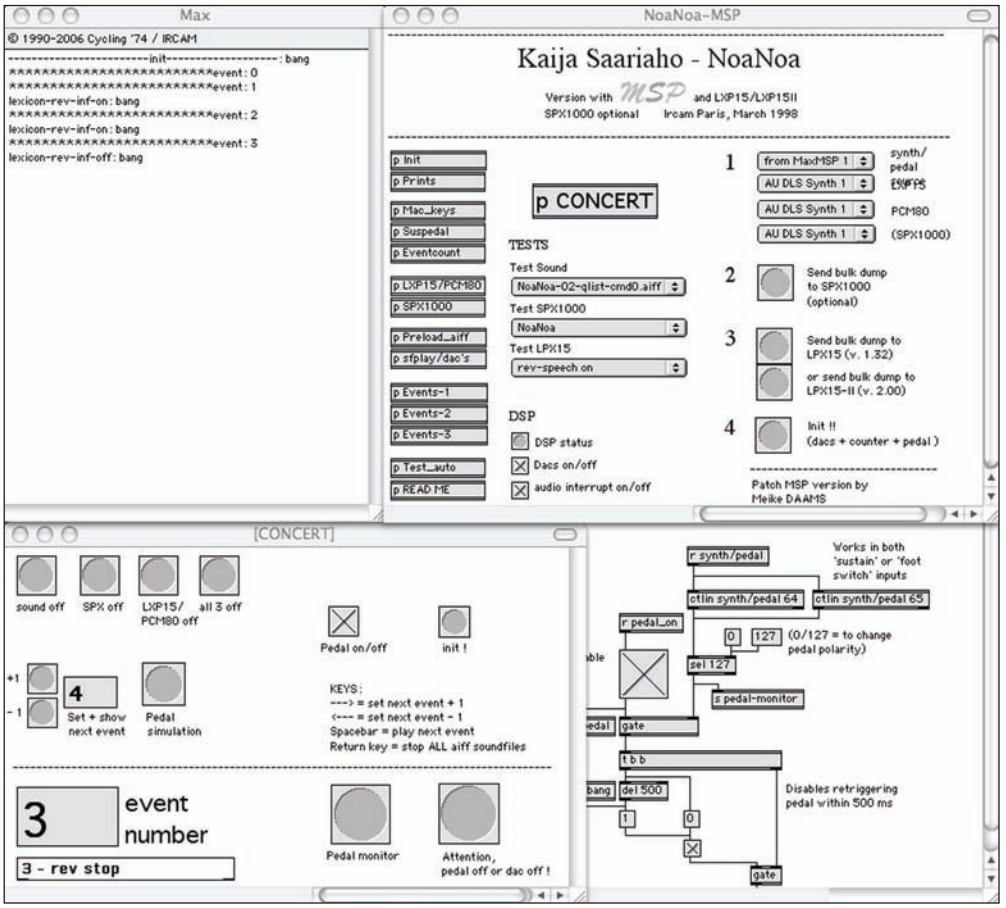
THE RELEVANCE OF ELECTROACOUSTIC PRACTICE

The development of electroacoustic music incorporating live instruments, over the last 15 years or so, provides an enlightening backdrop to the problem since this field presents us with a microcosm of our problem. It is also an area where a range of effective *musical* practices is beginning to be established and so offers glimpses into our future. Here only electronic or computer sound has to be delivered into a live setting. At a very basic level, adding video will simply be a matter of greater computing power, more disk space, and increased bandwidth, the principles remaining the same.

Without a doubt, the most significant development here has been the creation of flexible computer programming software environments such as MAX/MSP (Tools for New Media, n.d.), Pd (About Pure Data, n.d) and Supercollider (McCartney, n.d.). These provide the ability to process and/or create sound in almost any way imaginable. For example, live sound input from microphones can be processed in an incredibly wide variety of ways, an unlimited number of soundfiles/samples of practically any size can be played on cue, one can build synthesisers or samplers and incoming audio can be analysed and new musical material generated in response to it. The software can be controlled over MIDI (MIDI Manufacturers Association, n.d.) or Open Sound Control (Wright, 1997), high speed digital communications protocols, allowing use of a broad range of hardware controllers. Importantly, one can create a highly customised software configuration (known in MAX/MSP as a patch) for a particular piece, designed to work with a suitable hardware combination. One of the charms of these software packages is that they are absolutely open-ended, since you program from a very low level up. Although this can mean some hard work at the beginning, it has the considerable advantage that, unlike many music applications, no assumptions are made about what you want to achieve or how you might go about it, leaving you at the mercy of your imagination.

Kaija Saariaho's *Noa Noa* (1992) (Noa Noa, n.d.) for solo flute and "electronics" provides a very early but highly successful model of the sorts of interactive system one can build using this sort of

Figure 2. Computer (MAX/MSP) interface for Kaija Saariaho's Noa Noa. (1999, Chester Music Ltd. Used with permission.)



software. The original form of the piece requires a single computer running MAX/MSP for sample playback and two Lexicon PCM81 effects units controlled via MIDI, but in subsequent forms, the effects processing has been incorporated into the MAX/MSP patch to minimise the use of hardware. A microphone picks up the flute and passes the sound into the computer to be processed (via a multichannel audio interface); this is mixed with soundfile playback, and distributed over an eight-channel speaker system.

Control of the entire system is effectively in the hands of the flautist by means of a simple MIDI sustain pedal that needs to be depressed at cue points marked in the score. In the course of the roughly 10-minute piece, there are 64 cue points, and what happens at any particular depression of the pedal is entirely dependent on which cue number one is at. One of the 33 soundfiles might be played, the sound may move to a particular position in the eight-speaker array, some form of reverberation or harmonisation (pitch shifting) of the live flute sound may be applied, or perhaps a combination of these elements. To rehearse from anywhere in the score, one simply specifies the next cue point to be hit. This is all written into the patch in what amounts to an extra layer of the score.

The result is an incredibly beautiful, free and moving piece of music that always feels as if the flautist is in control of his or her expression. The way that the prerecorded soundfiles are used is

very interesting here. They are all short (ranging in length from 33 to less than 1 second), and in the parts where they are closely spaced, can overlap quite successfully (in the case of a fast performance). As a result, the flautist is always absolutely in control of the rhythm of any section. Many of the sounds are also derived from flute recordings, which creates a very close blend and dialogue with the playing. Combining this with the real-time treatments, it often becomes unclear whether the source is electronic or live. As a result, Saariaho's very careful conception and preparation of materials lets her create an integrated, expressive piece, and the listener hears something complex and vital that definitely belies the use of what might seem like a simple "on" button.

To some, the nature of the interaction between the flautist and computer (the pedal) may seem incredibly simplistic, given that it is now entirely possible to devise systems, such as gestural control, in which extra sensors applied to the instrument could be used to generate data that will control the computer (Diana Young's Hyperbow project for string instruments (Young, 2002) being a very interesting example of this). But it must also be remembered that purely in terms of flute technique, this is already a demanding piece, pushing the limits of what the instrument is capable of. So as far as the flautist is concerned, this simplicity is, in fact, a great strength. It lets them concentrate on their greatest skill, playing the flute beautifully. Many musicians actively dislike being "wired up to complicated systems" (to quote one I have worked with), despite having a real enthusiasm for working with new media. This is not being a Luddite, but rather a concern that their virtuosity and sound will suffer as a result of being distracted. If we want our work to be performed often by more than a just few specialists, we need to accommodate these concerns. Musicians need to enjoy their work to give of their best.

Saariaho's liking for simple interfaces can also be seen in her opera *L'Amour de Loin* (2000) (L'Amour de Loin, n.d.), where she uses a standard full-size MIDI keyboard to cue 88 eight-channel soundfiles. In the course of the entire opera, the pianist in the orchestra (in addition to their other duties playing a normal piano) simply plays a rising chromatic scale from the bottom to the top of the keyboard as indicated in the score- incredibly simple in performance and rehearsal, and incredibly effective.

More recently we have begun to see the development in electroacoustic music of a dedicated performer at the computer (or whichever interface is used), as an equal partner in the ensemble. This approach has much to recommend it. In my own *Xas-Orion* (2001) (New Ground, 2003) for oboe, cor-anglais, and computer, which again uses a MAX/MSP patch, the computer is performed directly from its alphanumeric keyboard using a "next cue" system, similar to that used in *Noa Noa*, combined with direct control of sound processing. So, for example, holding the letter D down will increase input of the oboe into the pitch shifting system, and holding F down will reduce it. This system works well with minimal hardware (just the laptop and PA), but for more intuitive and flexible control, these functions are also mapped to a hardware MIDI fader bank, in my case a Behringer B-Control BCF2000 (B-Control Fader BCF2000, n.d.), that connects directly to the laptop over USB.

In recent years, there has been an explosion in the use of systems like this, based as they are on relatively cheap, commercially available technology, and they represent a new type of musical instrument in their own right, which needs to be learnt just as one might learn to play the clarinet. This has developed naturally from recording studio technology, where, after all, most electroacoustic composers learn their craft; so these instruments usually combine familiar features from mixing desk and computer operation with the occasional use of MIDI keyboard and foot pedal controls. Subject to evolution over the years, these types of hardware control are not as arbitrary as

they may seem. They have proved themselves to be effective, providing (when well engineered) the sensitivity and good tactile feedback we find in other musical instruments. A less common, but highly effective device that can be added to this type of set up is Yamaha's excellent BC3A (Wind Guitar and Foot Controllers, n.d.), which generates MIDI from a small breath-sensitive mouthpiece. With a combination of these tools, we can easily provide enough simultaneous dimensions for control of the software, if this is intelligently laid out, always bearing in mind as we create the work how practically it is to be performed.

As works like *Noa Noa* show, the success or failure of these systems lies much more in how the software has been configured to work for, and as a part of, the piece in question than in the details of the hardware. Quality of content and flexibility of delivery can more than adequately compensate for simplicity of execution.

This instrumental approach also fits naturally into the world of music-making. Whether working from a score or improvising, traditional skills of ensemble playing, listening and response are still used in a way that other musicians can relate to easily and quickly.

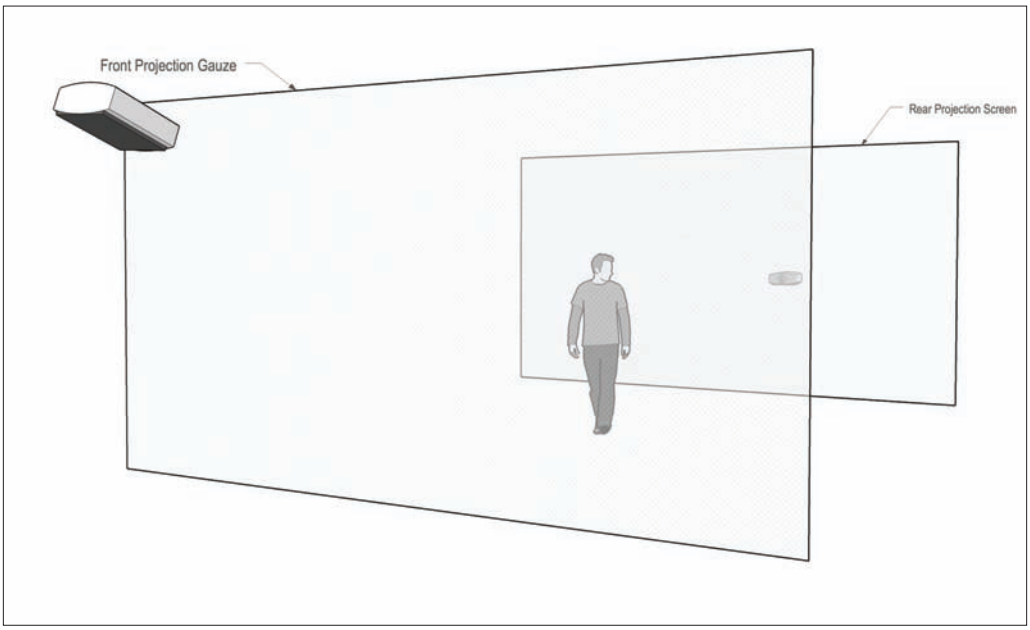
INCORPORATING VIDEO: REQUIREMENTS

Before we can design a system to deliver video elements for opera, we need to have a clear idea of what our ideal requirements are. I believe we should aim to produce works that are highly integrated and utilise the full range of possibilities offered by the moving image. As the most recent addition to an established art form, it is important that video never seems like a tacked-on afterthought, so we have to integrate on three levels: space (in terms of how video works on stage), narrative (content), and time (synchronisation).

The placing of video into the acting space needs careful consideration, and is a significant part of the initial design process. The mismatch between its essentially two-dimensional nature and the three dimensions of the theatre can be problematic, especially if there are significant interactions between what is live and onscreen. Multiple layers, involving several video projectors and the use of gauze screens, can provide a very good solution to this. Flat screens are, of course, not the only surfaces one can project on to, but the use of more interesting shapes may require some form of geometry correction so that the image does not become distorted undesirably (this is often required for situations where the projector is off axis to the screen anyway). Video projectors can do very simple geometry correction to remove keystoning, but anything more complex will have to be handled by computer software. It is also worth remembering that front projection directly onto the actors will throw shadows and can be rather blinding for them, possibly interfering with their ability to see the conductor. This can have disastrous consequences, so if space allows, rear projection is preferable. In my own experience, a combination of rear and front projection can be made to work very well. An additional consideration is whether the actors can actually see the video they are to interact with, and hidden monitors may have to be provided to assist with this.

Modern audiences are very fluent when it comes to receiving information through imagery, so we do not need to shy away from using video that contains important narrative strands, providing information that is essential to the dramatic development. For an effective blend, information content should be distributed evenly between all the elements of the opera-words, action, images, and sound-and this has to be written into the work, from the outset, with a clear understanding of why any particular piece of information is using a particular medium. Video allows us to incorporate text or graphics (moving or static), be used as an extension of lighting or set design, show our characters participating in events that

Figure 3. A simple, but highly flexible, stage layout allowing layering of imagery



do not take place on stage, or present them in other places or times. It enables presentation of multiple simultaneous viewpoints. An important part of this is that at different times in the work, we see the actors both live and on screen, and this could be achieved both by filming them beforehand or using live video from cameras that could be on or offstage. Using both, with the ability to composite these video streams live, can produce some startlingly effective results.

Video can be transformed in real time through the use of a very wide range of effects: adjusting colour, position, granulation, and so forth. This is not always as exciting as it may at first appear, and has to be used with caution, but is capable of adding meaning and expression in some very interesting ways.

Tight synchronisation between music, action, and video is essential for opera. Obviously, words and music are completely locked together in that they are sung, and this naturally leads on to synchronisation of action and music. This is a basic condition for this art form, and video must participate on equal terms. As we have already discussed, there is also a need for flexibility in tempo, leading to variability in the flow of time. So we need let the video follow the music (conductor or singer), standing the relationship we are familiar with, from film or TV, on its head. This requires a flexible and accurate cueing system; flexible enough to follow the score in real time and fast enough to respond instantaneously. Drawing again from the example of *Noa Noa*, the best approach would seem to be to use many short cues to piece the video together, but somehow achieve this without producing something choppy and fragmented. The rate at which we cut or dissolve from one image to another has great expressive effect, and is part of the language of film that we all understand, so this has to match the tone of the scene in question. Layers of video that can coexist would seem to be a solution to this problem. The overall effect required is deconstructed into smaller units that are then reassembled in real time. These can be divided over multiple screens or within one projection. If we are using

prerecorded clips, it is clear that they need to be of sufficient length to get from onset cue to offset cue at the slowest imaginable tempo (in other words, they should all be longer than is absolutely necessary). In terms of significant information content, these clips need to be "front weighted so that in a particularly fast performance, we do not cut away before the important event or image is shown. Another means of providing variable length for cues is being able to vary the frame rate in real time. This can be very effective, but is not appropriate for all types of material. For example, we are very sensitive to speed when viewing scenes of people in action, so a sense of slow-motion in this kind image may not be appropriate to the nature of the drama at that time. For film of clouds passing overhead, speed variation can be much less noticeable. Live video streams are, of course, inherently flexible, making them ideal in this situation, but there has to be a justification for using them rather than watching it happen on stage, and there are limitations on what you can set up. Again, a system that allows compositing in real time with other material would have great advantages. As we can see, the nature of the video material, and how it can be de- and reconstructed, is of great importance to the success of building a system that allows temporal flexibility, with content and form working together to achieve this.

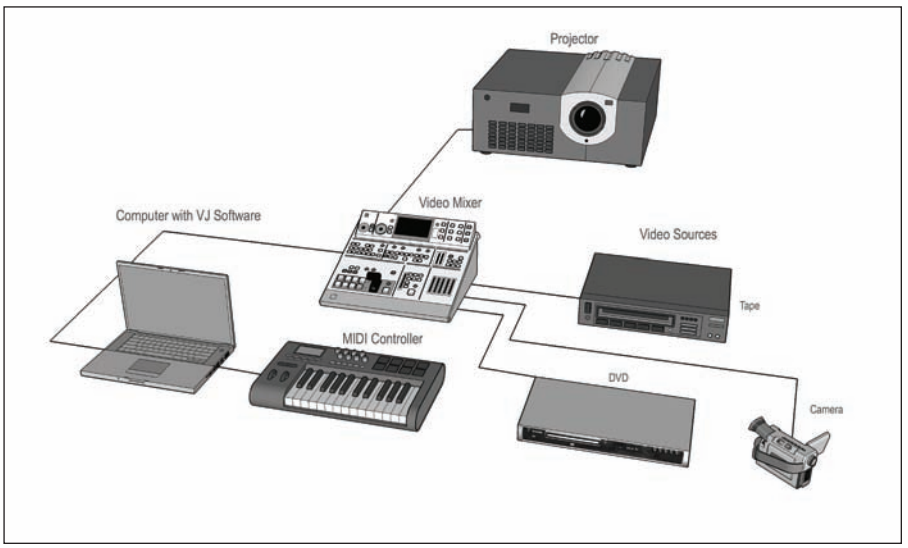
Clearly, the use of all these different layers will produce work of considerable complexity that will require careful preparation and planning. For *Black and Blue* (2004), we produced a cinema-style storyboarded script at a very early stage of the creative process, which proved immensely useful. This script was then regularly "cross-checked" with regards to practicality of means of production and design, and updated as our conception of the piece developed during writing and composition, with the rule that whenever we decided to include something, we had to be able to answer the question of how it was going to be done. Later, it became very easy to extract lists of sequences that needed to be filmed; images, graphics, and sounds to be collected; props required; and so forth, from this master script, as we progressed towards production.

Although based in the world of theatre rather than opera, *Super Vision* (2005) provides an inspiring example of what we can hope to achieve. Produced by the New York based Builders Association (The Builders Association, n.d.) and dbox (dbox dialog, strategy and design, 2005), the piece explores how we now live in a "post-private society, where personal electronic information is constantly collected and distributed. The data files collected on us circulate like extra bodies, and these 'data bodies' carry stains that are harder to clean than mud or sin" (Super Vision, n.d.). The piece effectively fulfils the requirements, outlined previously, on all three levels. The design of the stage space and use of multiple layers of projection is beautifully conceived, and the themes of surveillance, digital communication, and identity theft naturally lend themselves to, and provide opportunities for, the chosen media in compelling ways. Synchronisation is seamless and precise, responding organically to the pacing of the scenes. It is a short step from this kind of work to a form of opera that is equally satisfying.

INCORPORATING VIDEO: PRACTICE

Developments in computer software and hardware have begun to allow live and prerecorded video to be used with the same freedom that has been possible in the audio field. In fact, some of the same software that we have discussed earlier now comes with extensions for video processing and delivery. So, for example, MAX/MSP is now MAX/MSP/Jitter, with Jitter providing this extra functionality (Tools for New Media, n.d.). The fact that these extensions are part of an already well-proven programming environment, with the possibility of using the same control systems and

Figure 4. A typical analog VJ system

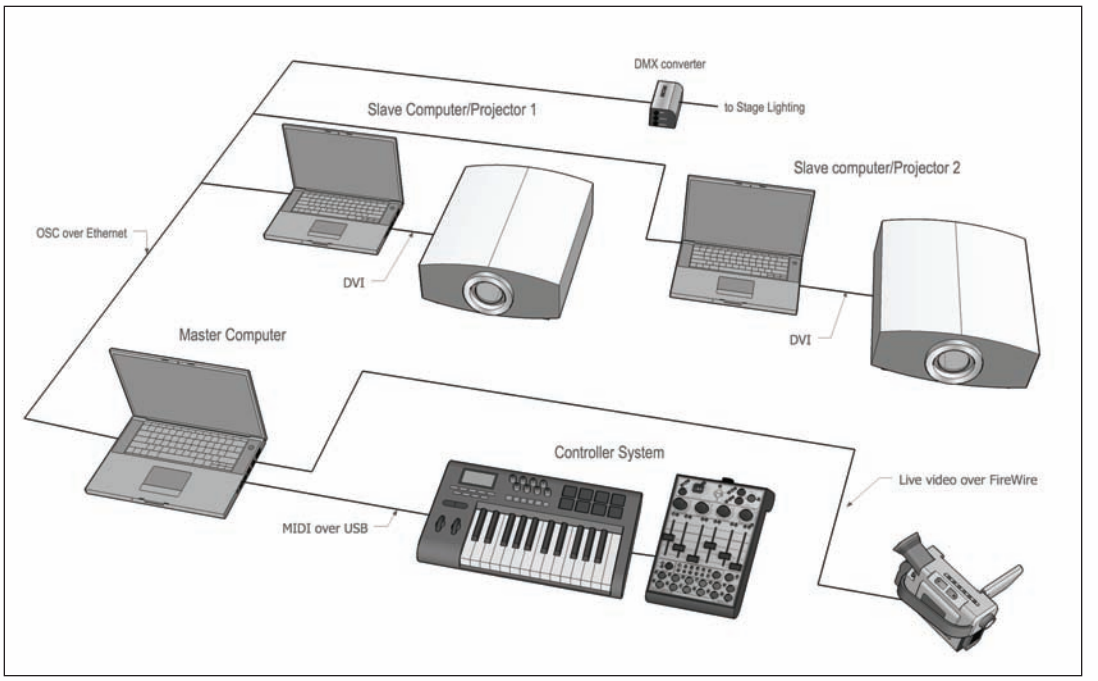


tight integration with sound, makes it immediately attractive for musical multimedia work. Other software more specifically specialised for video work, such as ArKaosVJ MIDI (Arkaos VJ 3.6 MIDI, 2006), Modul8 (Modul8, 2006), and Watchout (Dataton Watchout, n.d.), also offer similar abilities, each with its own particular emphasis.

Using the instrumental approach described earlier, we can perform video just as we can perform electronic music and, of course, VJs have been doing precisely this sort of thing for some years now, so we should expect to be able to derive systems from this field. VJing has its roots in clubbing, a world that is just as driven by the musical experience as opera, so many of the concerns outlined apply here too. What is significantly different to our concerns is a general lack of interest in narrative, which would have little place in a club anyway, so the tendency here is towards an improvisational graphic/abstract quality that suits the club experience. Much use is made of short looping sequences (in harmony with forms of dance music), with an emphasis on layering, compositing, and video effects. Larger scale media, which might carry dramatic content, are avoided. Given that VJing has always been at the forefront of technology, and has developed an aesthetic incredibly quickly, may also have a lot to do with technical limitations of systems in the past in terms of being able to deal with the very large file sizes.

The tools of the trade are usually a MIDI keyboard-triggering software on a laptop, some additional sources of video, and a vision mixer, using analog signal cabling to connect everything together. Substitute a sound mixer for the vision one and we can see the close connection to the standard electroacoustic set up. A fundamental principle of VJing is building up rich visual textures by layering smaller units. Software like Modul8 has this layering concept at the heart of its design. Here, up to five layers of video or still images are mixed and affected in real time. MIDI control allows accurate synchronised cueing of the entry and departure of these layers into the overall mix, and can control effects settings, playback speed, and layer transparency, among other things. Unfortunately, live video input from a camera does not appear to be possible. ArKaosVJ extends this by offering up to 50 layers and live input streams, and has a more intuitive MIDI interface.

Figure 5. A proposed MAX/MSP/Jitter based system—more computer/projector pairs can be added as necessary



A system like Watchout, which has its roots in corporate presentation rather than VJing, is interesting because of its high-quality support for multiple video projectors. This is achieved by using several computers networked over Ethernet. Each projector is driven by a dedicated computer, and these are all controlled from a separate production machine. The system is scaleable, can use multiple live input streams, and has some attractive features, including the ability to project over large areas using several aligned projectors that are "edge blended" to create a seamless single image, and support for advanced geometry correction to allow the use of curved screens. These make it very suitable for theatrical work, but, as yet, there is no support for MIDI or OSC control, which is a serious disadvantage, and the cueing system is not as intuitive as it could be. If we could combine Watchout's attractive features with the ease of use of something like ArKaosVJ and a custom cueing and control system, we would have something truly desirable for our purposes.

Of course, this is possible through the use of a programming environment rather than dedicated software, and MAX/MSP/Jitter provides us with precisely this. The Jitter extensions are highly optimised for video, enable any of the processes outlined, and because the audio and video system for a particular work can coexist in the same patch or in several patches running simultaneously on multiple computers linked via Ethernet and OSC, we can achieve the very high level of integration necessary for opera. Since MAX/MSP/Jitter can generate data as well as receive it, there is also the considerable advantage of being able to merge control of video and stage lighting. These obviously need to work together, and it is relatively straightforward to produce industry standard DMX protocol commands, to which any decent lighting board will respond.

It is clear to me that along the model of a VJ, we will need some form of dedicated performer to operate the "video instrument" under the direction (and so within sight) of the conductor.

Adding this workload to a pre-existing member of the production team or band is not a particularly desirable option, as they all have plenty to do already. A MIDI keyboard, combined with faders, foot pedals, and access to the computer is more than adequate as an interface if, as discussed before, the software is suitably designed, and we make sure that the computer is powerful enough to be instantaneously responsive. There ar,e of course, other possible interfaces. For example, one could be based on a percussion rather than keyboard model, but this one has precedence, and I like to think of it as a distant, highly developed offspring of the "Clavier à Lumière" Alexander Scriabin called for in his orchestral *Prometheus (The Poem of Fire)* as far back as 1911. Playing this "instrument" should feel familiar and easy to someone with experience as a VJ or electroacoustic musician, allowing a full range of performance styles from tightly specified writing through to free improvisation. It will be interesting to see how a new practice develops for this "instrument," just as MIDI keyboard technique has grown from previous forms of keyboard skills (piano and organ).

I am developing this approach for my current work, *The Girl Who Liked to be Thrown Around* (2006), my third collaboration with the writer Deepak Kalha. The piece is small in scale, takes the form of a psychological striptease, and is scored for a single woman, alto and bass flutes, and computers with two performers for sound and video, respectively. In contrast to our earlier work on *Black and Blue*, using a video performer in this has led us to think of many of the video elements in expressly musical terms-as notes to be composed and orchestrated–creating, if you will, visual melody, harmony, and counterpoint to match their counterparts in the story and music. This is not to say that the video is essentially abstract, as music is; far from it, in fact, with video having both diegetic and nondiegetic roles in the narrative. However, it does seem to promote a very nonlinear style of storytelling. This is also

Figure 6. A scene from The Girl Who Liked To Be Thrown Around*, using rear projection and a plasma screen on stage*

having interesting effects on other aspects, such as the dramaturgy and even how the words are put together, encouraging the use of multiple simultaneous layers of deconstructed material that conceal or reveal each other, depending on context.

Given this system of control for video, the directions for the performer might as well also take musical form, with events notated in a score. In the performance of opera, everyone is reading a score, from conductor to stage manager, and those who are not (the singer/actors) have memorised it. It would be simple and natural for the video performer to have a score too. It is worth remembering that the 20th century has seen considerable developments in the forms scores can take: completely standard notation, use of Lutoslawskian "mobiles," proportional notation, or the use of pure graphics provide a range of solutions to any imaginable performance situation that musicians are now quite familiar with, so this is not as limiting as it may seem.

Although it is technically possible to use computer-score-following technology to let us dispense with a performer (Philippe Manoury's *En Echo* (1993) (Manoury, 1993) for soprano and computer is an early and rather beautiful example from electroacoustic music), I completely fail to see what benefit this brings. The computer has to be operated by someone anyway, and in the context of opera, it seems reasonable to expect that person to have well-developed musical skills. More fundamentally, no matter how good the technology gets, the most interactive, intuitive thing I can think of is a person. Nothing is better at judging the energy of a scene in the moment. Human responses will make for humanity in performance, and when telling stories about people, this is vitally important.

Having said that a dedicated performer is required, putting some of the control into the hands of the singer/actors could have some very interesting results. One of the most appropriate ways of doing this would be through the use of motion tracking and gestural control. The EyesWeb software from InfoMus Lab, University of Genova (Camurri, Hashimoto, Richetti, Trocca, Suzuki, & Volpe, 2000) or Cyclops extensions to Jitter (Tools for New Media, n.d.) offer the ability to analyse incoming video in order to create control data that could be used in any number of ways, allowing the singers on stage to directly control events in the video or audio domain. The system has the advantage that it leaves the performer entirely free, there is no need to be wired up with special sensors, and we may well have live video feeds already in place.

A particularly elegant implementation of this is to bathe the stage in infrared light and then use a video camera, fitted with an infrared filter, as the input to the tracking system. This has the great advantage of working independently of any stage lighting, and the infrared light is, of course, invisible. Several actors can be tracked at once, and their interactions can be analysed to a remarkable degree, whether they are touching, moving towards or away from each other, behind or in front, and so forth. It is even possible to individually track different parts of the actor, the head, arms, and legs, and assign different triggering mechanisms to each part. As to the question of what these triggers might do, the possibilities are almost endless. We can control audio and video equally well, setting off sequences, letting imagery follow the actor around the space, defining points to cut from one scene to another, or altering the quality of the image or sound in a myriad of ways.

This can create some fascinating interactions between the live and video worlds, as has already been demonstrated in dance works by companies like Troika Ranch (Farley, 2002), but I do want to emphasise the use of the phrase *some of the control* here. Given the probably very large number of cue events an opera may contain, it would be unreasonable to expect to use this as the only means of control, and there are technical restraints on the amount of useful data that can be gener-

ated, since singers are rarely as mobile as dancers. Again, we need to be careful of overburdening our singers, and if we are striving for any form of naturalism in the acting, we will need to be subtle. But with careful planning, this could be used as an effective addition to what is already there. The great immediacy of response can create very striking effects that can be achieved in no other way, and this undoubtedly provides a very attractive way of enhancing the integration between action and video.

FUTURE TRENDS AND CONCLUSION

It is always surprising to me that opera, the form that always tried to claim the title of "total art work," has been so slow on the uptake when it comes to digital media. Undoubtedly this has much to do with the way culture of opera developed over the course of the 20th century, slowly turning it's gaze backward to the point where many contemporary composers refuse to use the term, preferring a range of inelegant alternatives (Lachenmann's "music with images,", Reich's "music video theatre," etc.). But I also believe that there has been a certain amount of waiting for the technology to become good enough to suit the form. Producing new opera is difficult and expensive, with the result that those who do it have to be pretty passionate about the form in the first place. This passion makes people very unwilling to dilute opera's natural strengths for the sake of including new media, even when there has been an understanding of, and enthusiasm for, the potential they offer. Very similar concerns apply in theatre, and here too, the amount of work of this type has been very small as a proportion of total output.

But really rather suddenly, the technology is ready, and not only ready, but easily available and relatively cheap. The rapid advance in computing power that we are experiencing will lead to big improvements in performance, ease of use, and stability very quickly, making the technology ever more attractive. There are, of course, a lot of new skills that need to be learnt, mainly in computing and programming, but precisely because these are computer based, there is instantly access to plenty of very active online communities where information and techniques are traded.

This is going to lead to an explosion of activity in this area, as people finally have the means to realise projects that have probably been in the back of their minds for a considerable amount of time. In a related sphere, the huge recent growth of VJing seems to bear this out. Eventually, we could see this becoming the dominant form for live dramatic performance-based art. The marriage of live and cinematic elements allowed by the incorporation of digital multimedia is irresistible, greatly expanding the narrative and dramatic possibilities. Even something as simple as the use of close-up, previously not possible on stage, can massively increase the power of a scene, if the acting is sufficiently good. With these elements meaningfully incorporated into the storytelling, we will be able to create sophisticated layered narratives that speak a language that modern, visually literate audiences have come to expect, and that have the power to engage and move those audiences. This need not be at the expense of the fundamentally musical nature of opera either. From this point of view, the form will not change at all. We can shape the use of multimedia to our musical requirements, rather than the other way round, so that one does not detract from the other, but rather strengthens what is already there.

Questions, of course, remain as to how successful the results of all this activity are, and what form the successful examples will take. The more work there is out there, in front of a critical public, and the greater the communication between practitioners, the better we will know the answers to this. Audiences will tell us what works and what does not. In the meantime, we need to develop effective working methods for

both creation and performance, drawing from as many fields as possible, to create a language for this new form of an old classic.

REFERENCES

About Pure Data. (n.d). Retrieved April 5, 2006 from http://puredata.info/

Arkaos VJ 3.6 MIDI. (2006). Retrieved April 5, 2006, from http://www.arkaos.net/software/vj_description.php

B-Control Fader BCF2000. (n.d.). Retrieved April 2, 2006, from http://www.behringer.com/BCF2000/index.cfm?lang=ENG

The Builders Association. (n.d.). Retrieved April 5, 2006, from http://www.thebuildersassociation.org/flash/flash.html?homepage

Camurri, A,. Hashimoto, S., Richetti, M., Trocca, R., Suzuki, K., & Volpe, G. (2000). EyesWeb—Toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal, 24*(1), 57-69.

Dataton Watchout. (n.d.). Retrieved April 3, 2006, from http://www.dataton.com/watchout

dbox dialog, strategy and design. (2005). Retrieved April 2, 2006, from http://www.dbox.com/

Farley, K. (2002). Digital dance theatre: The marriage of computers, choreography and techno/human reactivity *Body, Space and Technology.*

L'Amour de Loin. (n.d.). Retrieved April 3, 2006, from http://www.chesternovello.com

Manoury, P. (1993). *En Echo, the marriage of voice and electronics.* Retrieved April 5, 2006, from http://musicweb.koncon.nl/ircam/en/extending/enecho.html

McCartney, J. (n.d.). *Supercollider. A real time audio synthesis programming language.* Retrieved April 5, 2006, from http://www.audiosynth.com/

MIDI Manufacturers Association. (n.d.). Retrieved April 5, 2006, from http://www.midi.org/

Modul8. (2006). Retrieved April 3, 2006, from http://www.garagecube.com/modul8/index.php

New Ground. (2003). Retrieved April 6, 2006, from http://www.oboeclassics.com/NewGround.htm

Noa Noa. (n.d.). Retrieved April 3, 2006, from http://www.chesternovello.com

Oliva, M. (2004). Retrieved March 19, 2006, from http://www.madestrange.net/black.htm

Reich, S., & Korot, B. interviewed by David Allenby. (2001). Retrieved April 5, 2006, from http://www.boosey.com/pages/opera/OperaNews.asp?NewsID=10260&MusicID=15153

Richmond, J. (1989). *Valis points to exciting possibilities for growth of opera.* Retrieved from http://www-tech.mit.edu/V109/N28/valis.28a.html

Super Vision. (n.d.). Retrieved April 5, 2006, from http://www.superv.org/

Tools for new Media. (n.d.). Retrieved April 5, 2006 from http://www.cycling74.com/

Truax, B. (1996). Sounds and sources in *Powers of Two*: Towards a contemporary myth. *Organised Sound, 1*(1), 13-21.

Weiss, N. (1966). Film and Lulu. *Opera, 17*(9), 708.

Wind Guitar and Foot Controllers. (n.d.). Retrieved April 6, 2006, from http://www.yamaha.com/yamahavgn/CDA/ContentDetail/ModelSeriesDetail/0,,CNTID%253D1321%2526CTID%253D,00.html

Wright, M., & Freed, A. (1997). *Open sound control: A new protocol for communicating with sound synthesizers.* Paper presented at the International Computer Music Conference 1997, Thessaloniki, Greece.

Young, D. (2002). *The Hyperbow. A precision violin interface*. Paper presented at the International Computer Music conference 2002, Gothenburg, Sweden.

This work was previously published in Interactive Multimedia Music Technologies, edited by K. Ng and P. Nesi, pp. 151-166, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.11
Face Animation:
A Case Study for Multimedia Modeling and Specification Languages

Ali Arya
University of British Columbia, Canada

Babak Hamidzadeh
University of British Columbia, Canada

ABSTRACT

This chapter will discuss the multimedia modeling and specification methods, especially in the context of face animation. Personalized Face Animation is and/or can be a major user interface component in modern multimedia systems. After reviewing the related works in this area, we present the ShowFace streaming structure. This structure is based on most widely accepted industry standards in multimedia presentations like MPEG-4 and SMIL and extends them by providing a higher level Face Modeling Language (FML) for modeling and control purposes and by defining image transformations required for certain facial movements. ShowFace establishes a comprehensive framework for face animation consisting of components for parsing the input script, generating and splitting the audio and video "behaviors," creating the required images and sounds, and eventually displaying or writing the data to files. This component-based design and scripted behavior make the framework suitable for many purposes including web-based applications.

INTRODUCTION

Specifying the components of a multimedia presentation and their spatial/temporal relations are among basic tasks in multimedia systems. They are necessary when a client asks for a certain presentation to be designed, when a media player receives input to play, and even when a search is done to retrieve an existing multimedia file. In all these cases, the description of the contents can include raw multimedia data (video, audio, etc.) and textual commands and information. Such a description works as a Generalized Encoding, since it represents the multimedia content in a form not necessarily the same as the playback format,

and is usually more efficient and compact. For instance a textual description of a scene can be a very effective "encoded" version of a multimedia presentation that will be "decoded" by the media player when it recreates the scene.

Face Animation, as a special type of multimedia presentation, has been a challenging subject for many researchers. Advances in computer hardware and software, and also new web-based applications, have helped intensify these research activities, recently. Video conferencing and online services provided by human characters are good examples of the applications using face animation. Personalized Face Animation includes all the information and activities required to create a multimedia presentation resembling a specific person. The input to such a system can be a combination of audio/visual data and textual commands and descriptions. A successful face animation system needs to have efficient yet powerful solutions for providing and displaying the content, i.e., a content description format, decoding algorithms, and finally an architecture to put different components together in a flexible way.

Although new streaming technologies allow real-time download/playback of audio/video data, bandwidth limitation and it's efficient usage still are, and probably will be, major issues. This makes a textual description of multimedia presentation (e.g., facial actions) a very effective coding/compression mechanism, provided the visual effects of these actions can be recreated with a minimum acceptable quality. Based on this idea, in face animation, some researches have been done to translate certain facial actions into a predefined set of "codes." Facial Action Coding System (Ekman & Friesen, 1978) is probably the first successful attempt in this regard. More recently, MPEG-4 standard (Battista, et al., 1999) has defined Face Animation Parameters to encode low-level facial actions like jaw-down, and higher level, more complicated ones like smile.

Efficient use of bandwidth is not the only advantage of multimedia content specifications like facial action coding. In many cases, the "real" multimedia data does not exist at all and has to be created based on a description of desired actions. This leads to the whole new idea of representing the spatial and temporal relation of the facial actions. In a generalized view, such a description of facial presentation should provide a hierarchical structure with elements ranging from low-level "images," to simple "moves," to more complicated "actions," to complete "stories." We call this a Structured Content Description, which also requires means of defining capabilities, behavioural templates, dynamic contents, and event/user interaction. Needless to say, compatibility with existing multimedia and web technologies is another fundamental requirement, in this regard.

Having a powerful description and specification mechanism, also is obviously powerful in search applications that currently suffer when looking for multimedia content. MPEG-7 standard (Nack & Lindsay, 1999) is the newest arrival in the group of research projects aiming at a better multimedia retrieval mechanism.

Considering three major issues of Content Delivery, Content Creation, and Content Description, the following features can be assumed as important requirements in a multimedia presentation systems (Arya & Hamidzadeh, 2002):

1. Streaming, i.e., continuously receiving/displaying data
2. Structured Content Description, i.e., a hierarchical way to provide information about the required content from high-level scene description to low-level moves, images, and sounds
3. Content Creation (Generalized Decoding), i.e., creating the displayable content based on the input. This can be decoding a compressed image or making new content based on the provided textual description.
4. Component-based Architecture, i.e., the flexibility to rearrange the system components, and use new ones as long as a certain interface is supported.

5. Compatibility, i.e., the ability to use and work with widely accepted industry standards in multimedia systems.
6. Minimized Database of audio/visual footage.

The technological advances in multimedia systems, speech/image processing, and computer graphics, and also new applications especially in computer-based games, telecommunication, and online services, have resulted in a rapidly growing number of publications regarding these issues. These research achievements, although very successful in their objectives, mostly address a limited subset of the above requirements. A comprehensive framework for face animation is still in conceptual stages.

The *ShowFace* system, discussed later, is a step toward such a framework. It is based on a modular structure that allows multimedia streaming using existing technologies and standards like MPEG-4, Windows Media and DirectX/DirectShow (http://www.microsoft.com/windows/directx), and XML (http://www.w3.org/XML). The components independently read and parse a textual input, create audio and video data, and mix them together. Each component can be replaced and upgraded as long as it conforms to the ShowFace Application Programming Interface (SF-API). SF-API also allows other programs like web browsers to connect to ShowFace components directly or through a wrapper object called *ShowFacePlayer*. The system uses a language specifically designed for face animation applications. Face Modeling Language (FML) is an XML-based structured content description language that describes a face animation in a hierarchical way (from high-level stories to low-level moves), giving maximum flexibility to the content designers. Receiving FML scripts as input, *ShowFace* generates the required frames based on a limited number of images and a set of pre-learned transformations. This minimizes the image database and computational complexity, which are issues in existing approaches, as reviewed later.

In Section 2, some of the related works will be briefly reviewed. This includes different approaches to multimedia modeling and specification (content description in general), multimedia systems architectures to support those specification mechanisms, and eventually, content creation methods used in related multimedia systems. The basic concepts and structure of the *ShowFace* system will be discussed in Section 3 to 5. This includes the proposed Face Modeling Language (FML) and the system structure and components for parsing the input and creating the animation. Some experimental results and conclusions will be the topics of Sections 6 and 7, respectively.

RELATED WORK

Multimedia Content Description

The diverse set of works in multimedia content description involves methods for describing the components of a multimedia presentation and their spatial and temporal relations. Historically, one of the first technical achievements in this regard was related to video editing where temporal positioning of video elements is necessary. The SMPTE (Society of Motion Picture and Television Engineers) time coding (Ankeney, 1995; Little, 1994) that precisely specifies the location of audio/video events down to the frame level is base for EDL (Edit Decision List) (Ankeney, 1995; Little, 1994) that relates pieces of recorded audio/video for editing. Electronic Program Guides (EPGs) are another example of content description for movies in the form of textual information added to the multimedia stream.

More recent efforts by SMPTE are focused on Metadata Dictionary, which targets the definition of the metadata description of the content (see http://www.smpte-ra.org/mdd). These metadata can include items from title to subject and components. The concept of metadata description is a base for other similar researches like Dublin Core

(http://dublincore.org), EBU P/Meta (http://www.ebu.ch/pmc_meta.html), and TV Anytime (http://www.tv-anytime.org). Motion Picture Expert Group is also another major player in standards for multimedia content description and delivery. MPEG-4 standard, which comes after MPEG-1 and MPEG-2, is one of the first comprehensive attempts to define the multimedia stream in terms of its forming components (objects like audio, foreground figure, and background image). Users of MPEG-4 systems can use Object Content Information (OCI) to send textual information about these objects.

A more promising approach in content description is MPEG-7 standard. MPEG-7 is mainly motivated by the need for a better more powerful search mechanism for multimedia content over the Internet but can be used in a variety of other applications including multimedia authoring. The standard extends OCI and consists of a set of descriptors for multimedia features (similar to metadata in other works), schemes that show the structure of the descriptors, and an XML-based description/schema definition language.

Most of these methods are not aimed at and customized for a certain type of multimedia stream or object. This may result in a wider range of applications but limits the capabilities for some frequently used subjects like human faces. To address this issue MPEG-4 includes Face Definition Parameters (FDPs) and Face Animation Parameters (FAPs). FDPs define a face by giving measures for its major parts, as shown in Figure 1. FAPs on the other hand, encode the movements of these facial features. Together they allow a receiver system to create a face (using any graphic method) and animate based on low-level commands in FAPs. The concept of FAP can be considered a practical extension of Facial Action Coding System (FACS) used earlier to code different movements of facial features for certain expressions and actions.

After a series of efforts to model temporal events in multimedia streams (Hirzalla et al., 1995), an important progress in multimedia content description is Synchronized Multimedia Integration Language (SMIL) (Bulterman, 2001), an XML-based language designed to specify the temporal relation of the components of a multimedia presentation, especially in web applications. SMIL can be used quite suitably with MPEG-4 object-based streams.

There have also been different languages in the fields of virtual reality and computer graphics for modeling computer-generated scenes. Examples are Virtual Reality Modeling Language (VRML, http://www.vrml.org) and programming libraries like OpenGL (http://www.opengl.org).

Another important group of related works are behavioural modeling languages and tools for virtual agents. BEAT (Cassell et al., 2001) is an XML-based system, specifically designed for human animation purposes. It is a toolkit for automatically suggesting expressions and gestures, based on a given text to be spoken. BEAT uses a knowledge base and rule set and provides synchronization data for facial activities, all in XML format. This enables the system to use standard XML parsing and scripting capabilities. Although BEAT is not a general content description tool, it demonstrates some of the advantages of XML-based approaches. Other scripting and behavioural modeling languages for virtual humans are considered by other researchers as well (Funge et al., 1999; Kallmann & Thalmann, 1999; Lee et al., 1999). These languages are usually simple macros for simplifying the animation, or new languages, which are not using existing multimedia technologies. Most of the time they are not specifically designed for face animation.

Multimedia Content Creation

In addition to traditional recording, mixing, and editing techniques in the film industry, computer graphics research has been long involved in multimedia, especially in image/video creation. Two main categories can be seen in these works: 3-D

geometrical models (Blanz & Vetter, 1999; Lee et al., 1999; Pighin et al., 1998) and 2-D image-based methods (Arya & Hamidzadeh, 2002; Bregler et al., 1997; Ezzat & Poggio, 1998; Graf et al., 2000).

The former group involves describing the scene using 3-D data (as in VRML and OpenGL) and then rendering the scene (or sequence of frames) from any point of view. These techniques need usually complicated 3-D models, data and computation, but are very powerful in creating any view provided the model/data is inclusive enough. Due to the way images are generated and their inability to include all the details, most of these methods do not have a very realistic output. In the case of face animation, 3-D techniques are used by many researchers (Blanz & Vetter, 1999; Lee et al., 1999; Pighin et al., 1998). To reduce the size of required data for model generation, some approaches are proposed to create 3-D models based on a few (two orthogonal) 2-D pictures (Lee et al., 1999).

It is shown that any view of a 3-D scene can be generated from a combination of a set of 2-D views of that scene or by applying some transformations on them (Arya & Hamidzadeh, 2002; Ezzat & Poggio, 1998). This fact is base for the latter group of techniques, i.e., 2-D image-based. In a talking head application, Ezzat et al. (Ezzat & Poggio, 1998) use view morphing between pre-recorded visemes (facial views when pronouncing different phonemes) to create a video corresponding to any speech. Optical flow computation is used to find corresponding points in two images, solving the correspondence problem for morphing. Bregler et al. (1997) combine a new image with parts of existing footage (mouth and jaw) to create new talking views. Both these approaches are limited to a certain view where the recordings have been made. No transformation is proposed to make a talking view after some new movements of the head. In a more recent work based on MikeTalk, (Graf et al., 2000) recording of all visemes in a range of possible views is proposed, so after detecting the view (pose) proper visemes will be used. This way talking heads in different views can be animated but the method requires a considerably large database.

Defining general image transformations for each facial action and using facial feature points to control the mapping seem to be helpful in image-based techniques. TalkingFace (Arya & Hamidzadeh, 2002) combines optical flow and facial feature detection to overcome these issues. It can learn certain image transformations needed for talking (and potentially expressions and head movements) and apply them to any given image. Tiddeman et al. (2001) show how such image transformations can be extended to include even facial texture.

Multimedia Systems Architectures

Different architectures are proposed for multimedia systems. They try to address different aspects of multimedia, mostly streaming and playback. The main streaming systems, aimed at web-based transmission and playback, are Microsoft Windows Media, Apple QuickTime, and Real Networks RealVideo (Lawton, 2000).

Different architectures are also proposed to perform facial animation, especially as an MPEG-4 decoder/player (Pandzic, 2001). Although they try to use platform-independent and/or standard technologies (e.g., Java and VRML), they are usually limited to certain face models and lack a component-based and extensible structure and do not propose any content description mechanism more than standard MPEG-4 parameters.

STRUCTURED CONTENT DESCRIPTION IN *SHOWFACE*

Design Ideas

Describing the contents of a multimedia presentation is a basic task in multimedia systems.

It is necessary when a client asks for a certain presentation to be designed, when a media player receives input to play, and even when a search is done to retrieve an existing multimedia file. In all these cases, the description can include raw multimedia data (video, audio, etc.) and textual commands and information. Such a description works as a Generalized Encoding, since it represents the multimedia content in a form not necessarily the same as the playback format and is usually more efficient and compact. For instance a textual description of a scene can be a very effective "encoded" version of a multimedia presentation that will be "decoded" by the media player when it recreates the scene.

Although new streaming technologies allow real-time download/playback of audio/video data, bandwidth limitation and it's efficient usage still are, and probably will be, major issues. This makes a textual description of multimedia presentations (in our case facial actions) a very effective coding/compression mechanism, provided the visual effects can be recreated with minimum acceptable quality.

Efficient use of bandwidth is not the only advantage of facial action coding. In many cases, the "real" multimedia data does not exist at all and has to be created based on a description of desired actions. This leads to the whole new idea of representing the spatial and temporal relation of the facial actions. In a generalized view, such a description of facial presentation should provide a hierarchical structure with elements ranging from low-level "images," to simple "moves," to more complicated "actions," to complete "stories." We call this a Structured Content Description, which also requires means of defining capabilities, behavioural templates, dynamic contents, and event/user interaction. Needless to say, compatibility with existing multimedia and web technologies is another fundamental requirement, in this regard.

Face Modeling Language (FML) is a Structured Content Description mechanism based on eXtensible Markup Language. The main ideas behind FML are:

- Hierarchical representation of face animation
- Timeline definition of the relation between facial actions and external events
- Defining capabilities and behaviour templates
- Compatibility with MPEG-4 FAPs
- Compatibility with XML and related web technologies

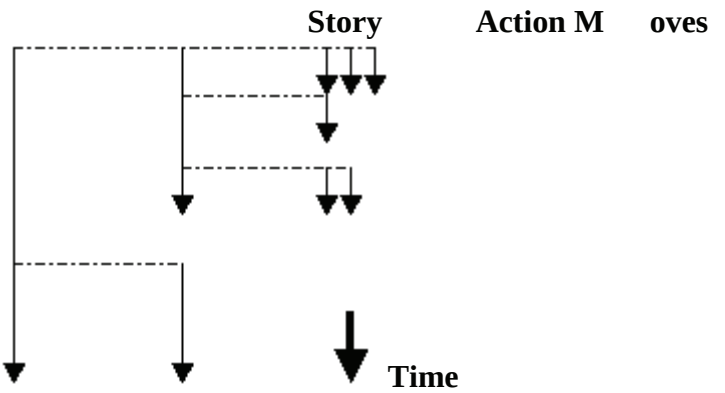
FACS and MPEG-4 FAPs provide the means of describing low-level face actions but they do not cover temporal relations and higher-level structures. Languages like SMIL do this in a general purpose form for any multimedia presentation and are not customized for specific applications like face animation. A language bringing the best of these two together, customized for face animation, seems to be an important requirement. FML is designed to do so.

Fundamental to FML is the idea of Structured Content Description. It means a hierarchical view of face animation capable of representing simple individually-meaningless moves to complicated high-level stories. This hierarchy can be thought of as consisting of the following levels (bottom-up):

- Frame, a single image showing a snapshot of the face (Naturally, may not be accompanied by speech)
- Move, a set of frames representing a linear transition between two frames (e.g., making a smile)
- Action, a "meaningful" combination of moves
- Story, a stand-alone piece of face animation

The boundaries between these levels are not rigid and well defined. Due to the complicated

Figure 1. FML timeline and temporal relation of face activities



and highly expressive nature of facial activities, a single move can make a simple yet meaningful story (e.g., an expression). The levels are basically required by the content designer in order to:

- Organize the content
- Define temporal relation between activities
- Develop behavioural templates, based on his/her presentation purposes and structure.

FML defines a timeline of events (Figure 1) including head movements, speech, and facial expressions, and their combinations. Since a face animation might be used in an interactive environment, such a timeline may be altered/determined by a user. So another functionality of FML is to allow *user interaction* and in general *event handling* (Notice that user input can be considered a special case of *external event*). This event handling may be in the form of:

- Decision Making; choosing to go through one of possible paths in the story
- Dynamic Generation; creating a new set of actions to follow

A major concern in designing FML is compatibility with existing standards and languages. Growing acceptance of MPEG-4 standard makes it necessary to design FML in a way it can be translated to/from a set of FAPs. Also, due to the similarity of concepts, it is desirable to use SMIL syntax and constructs, as much as possible.

Primary Language Constructs

FML is an XML-based language. The choice of XML as the base for FML is based on its capabilities as a markup language, growing acceptance, and available system support in different platforms. Figure 2 shows the typical structure of an FML.

An FML document consists, at higher level, of two types of elements: **model** and **story**. A **model** element is used for defining face capabilities, parameters, and initial configuration. A **story** element, on the other hand, describes the timeline of events in face animation. It is possible to have more than one of each element but due to possible sequential execution of animation in streaming applications, a **model** element affects only those parts of a document coming after it.

A face animation timeline consists of facial activities and their temporal relations. These activities are themselves sets of simple moves. The timeline is primarily created using two time container elements, **seq** and **par** representing sequential and parallel move-sets. A **story** itself

Figure 2. FML document map—Time-container and move-set will be replaced by FML time container elements and sets of possible activities, respectively

```
<fml>
    <model>        <!-- Model Info -->
         <model-info />
    </model>
    <story>        <!-- Story Timeline -->
         <action>
              <time-container>
                   <move-set>
                        < . . . >
                   <move-set>
                   < . . . >
              </time-container>
              < . . . >
         </action>
         < . . . >
    </story>
</fml>
```

Figure 3. FML primary time container

```
<seq begin="0">
    <talk begin="0">Hello World</talk>
    <hdmv begin="0" end="5" type="0" val="30" />
</seq>
<par begin="0">
    <talk begin="1">Hello World</talk>
    <exp begin="0" end="3" type="3" val="50" />
</par>
```

is a special case of sequential time container. The begin times of activities inside a **seq** and **par** are relative to previous activity and container begin time, respectively.

FML supports three basic face activities: talking, expressions, and 3-D head movements. They can be a simple move (like an expression) or more complicated (like a piece of speech). Combined in time containers, they create FML Actions. This combination can be done using nested containers, as shown in Figure 4.

FML also provides the means for creating a behavioral model for the face animation. At this time, it is limited to initialization data such as a range of possible movements and image/sound databases, and simple behavioral templates (subroutines). But, it can be extended to include behavioral rules and knowledge bases, especially for interactive applications. A typical **model** element is illustrated in Figure 5, defining a behavioral template used later in **story**.

Event Handling and Decision Making

Dynamic interactive applications require the FML document to be able to make decisions, i.e., to follow different paths based on certain events. To accomplish this **excl** time container

Figure 4. Nested time container

```
<action>
<par begin="0">
<seq begin="0">
        <talk begin="0">Hello World</talk>
   <hdmv begin="0" end="5" type="0" val="30" />
</seq>
   <exp begin="0" end="3" type="3" val="50" />
</par>
</action>
```

Figure 5. FML model and templates

```
<model>
    <img src="me.jpg" />
    <range dir="0" val="60" />
    <template name="hello" >
<seq begin="0">
    <talk begin="0">Hello</talk>
        <hdmv begin="0" end="5" dir="0" val="30" />
</seq>
    </template>
</model>
<story>
    <behavior template="hello" />
</story>
```

and **event** element are added. An event represents any external data, e.g., the value of a user selection. The new time container associates with an event and allows waiting until the event has one of the given values, then it continues with action corresponding to that value.

Compatibility

The XML-based nature of this language allows the FML documents to be embedded in web pages. Normal XML parsers can extract data and use them as input to an FML-enabled player, through simple scripting. Such a script can also use XML Document Object Model (DOM) to modify the FML document, e.g., adding certain activities based on user input. This compatibility with web browsing environments, gives another level of interactivity and dynamic operation to the FML-based system, as illustrated later in this chapter.

Another aspect of FML is its compatibility with MPEG-4 face definition/animation parameters. This has been achieved by:

- Translation of FML documents to MPEG-4 codes by the media player.
- Embedded MPEG-4 elements (**fap** element is considered to allow direct embedding of FAPs in FML document).

Figure 6. FML Decision Making and Event Handling

```
<event id="user" val="-1" />
<excl ev_id="user">
    <talk ev_val="0">Hello</talk>
    <talk ev_val="1">Bye</talk>
</excl>
```

Case Studies

Static Document

The first case is a simple FML document without any need for user interaction. There is one unique path the animation follows. The interesting point in this basic example is the use of loop structures, using **repeat** attributes included in any activity.

The **event** element specifies any external entity whose value can change. The default value for **repeat** is 1. If there is a numerical value, it will be used. Otherwise, it must be an **event** id, in which case the value of that **event** at the time of execution of related activity will be used. An FML-compatible player should provide means of setting external events values. *ShowFacePlayer* has a method called *SetFaceEvent*, which can be called by the owner of a player object to simulate external events.

Event Handling

The second example shows how to define an external event, wait for a change in its value, and perform certain activities based on the value. An external event corresponding to an interactive user selection is defined first. It is initialized to –1 that specifies an invalid value. Then, an **excl** time container, including required activities for possible user selections, is associated with the event. The **excl** element will wait for a valid value of the event. This is equivalent to a pause in face animation until a user selection is done.

It should be noted that an FML-based system usually consists of three parts:

- FML Document
- FML-compatible Player
- Owner Application

In a simple example like this, it could be easier to simply implement the "story" in the owner application and send simpler commands to a player just to create the specified content (e.g., face saying Hello). But in more complicated cases, the owner application may be unaware of desired stories or unable to implement them. In those cases, e.g., interactive environments, the owner only simulates the external parameters.

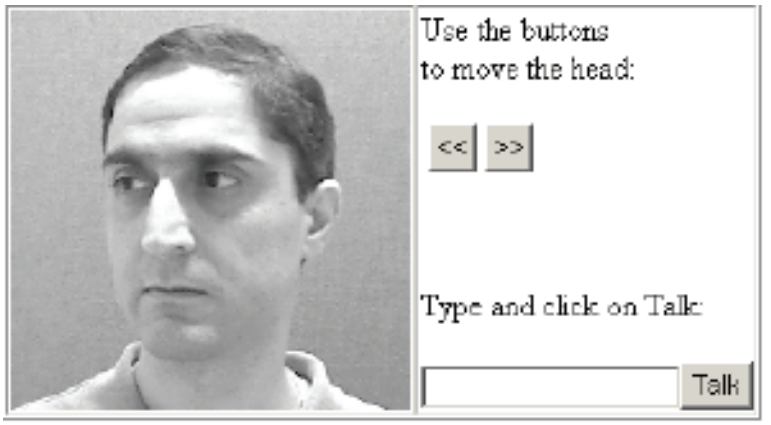
Dynamic Content Generation

The last FML example to be presented illustrates the use of XML Document Object Model (DOM)

Figure 7. Repeated activity—Using event is not necessary

```
<event id="select" val="2" />
< . . . >
<seq repeat="select">
    <talk begin="0">Hello World</talk>
<exp begin="0" end="3" type="3" val="50" />
</seq>
```

Figure 8. Dynamic FML generation



to dynamically modify the FML document and generate new animation activities.

The simplified and partial JavaScript code for the web page shown in Figure 8 looks like this:

```
function onRight()
{
    var fml = fmldoc.documentElement;
    var new=fmldoc.createElement("hdmv");
    new.setAttribute("dir","0");
    new.setAttribute("val","30");
    fml.appendChild(new);
}
```

More complicated scenarios can be considered, using this dynamic FML generation, for instance, having a form-based web page and asking for user input on a desired behavior and using templates in **model** section of FML.

CONTENT CREATION IN *SHOWFACE*

Feature-Based Image Transformation (FIX)

The FML parser component of *ShowFace* system determines the visual activities required in the face. These activities are transitions between certain face *states* like a viseme or expression. In a training phase, a set of image-based transformations is learned by the system, which can map between these face states. Transformations are found by tracking facial features when the model is performing the related transitions, and then applied to a given image, as illustrated in Figure 10. A library of transformations is created based on the following facial states:

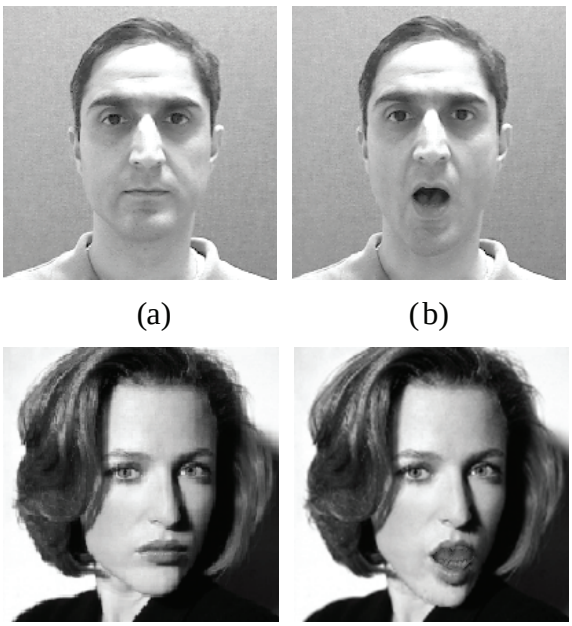
Figure 9. JavaScript code for FML event shown in Figure 6

```
function onLoad()
{
    facePlayer.ReadFML("test.fml");
    facePlayer.Run();
}

function onHelloButton()
{
    facePlayer.SetFaceEvent(
        "user", 0);
}

function onByeButton()
{
    facePlayer.SetFaceEvent(
        "user", 1);
}
```

Figure 10. Facial features and image transformations — (a) Model state 1, (b) Model state 2, (c) Target character in state 1, (d) Target character in state 2



- Visemes in full-view
- Facial expressions in full-view
- Head movements

For group 1 and 2, mappings for all the transitions between a non-talking neutral face and any group member are stored. In group 3 this is done for transitions between any two neighbouring states (Figure 11).

Each transformation is defined in the form of $T = (F,M)$ where T is the transformation, F is the feature set in the source image, and M is the mapping value for features. Source image information is saved to enable scaling and calibration, which is explained later. The feature set for each image includes face boundary, eyes and eyebrows, nose, ears, and lips. These feature lines and the facial regions created by them are shown in Figure 12.

The solid lines are feature lines surrounding feature regions, while dashed lines define face patches. The patches are defined in order to allow

Figure 11. Moving head states—The same three states exist for rotating to the left, in addition to a full-view image at the center

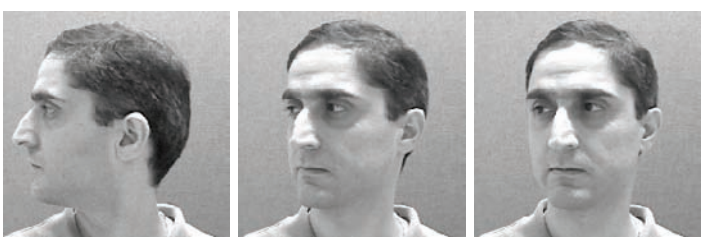


Figure 12. Facial regions are defined as areas surrounded by two facial feature lines, e.g., inside eyes or between lower lip and jaw (some face patches are removed (b) for simplicity.)

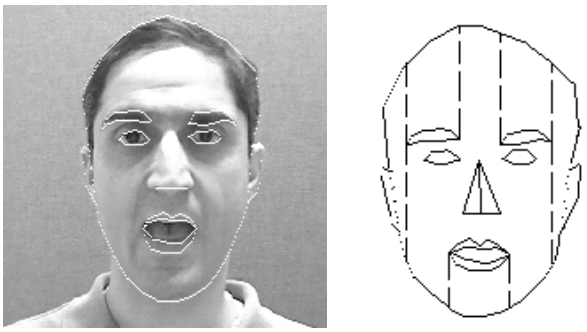
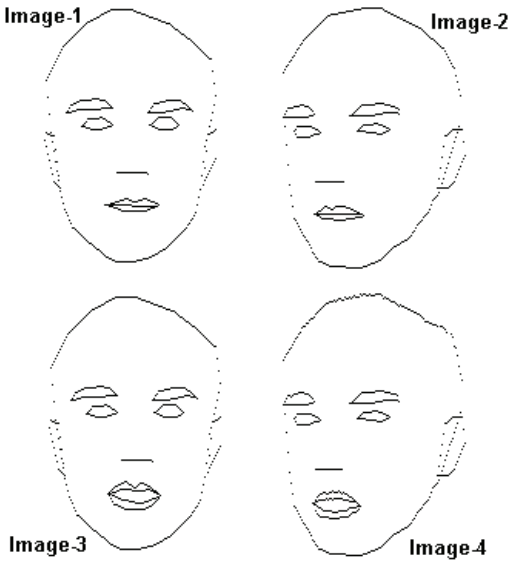


Figure 13. Feature transformation—Tij is the transformation between image-i and image-j



different areas of the face to be treated differently. Covisibility is the main concern when defining these face patches. Points in each patch will be mapped to points in the corresponding patch of the target image, if visible.

The transformations are done by first applying the mapping vectors for the feature points. This is shown in Figure 13. Simple transformations are those which have already been learned, e.g., T_{12} and T_{13} (assuming we only have *Image-1*). Combined transformations are necessary in cases when the target image is required to have the effect of two facial state transitions at the same time, e.g., T_{14}.

Due to the non-orthographic nature of some head movements, combined transformations involving 3-D head rotation cannot be considered a linear combination of some known transfor-

mations. Feature mapping vectors for talking and expressions (which are learned from frontal view images) need to be modified when applied to "moved" heads.

$$T_{14} = a\ T_{12} + b\ T'_{13}$$
$$T'_{13} = f_p(T_{12}, , T_{13}) = T_{24}$$

where f_p is Perspective Calibration Function and *a* and *b* are coefficients between 0 and 1 to control transition progress. T_{13} will also be scaled based on face dimensions in source/target images.

When the *Image-2* is given, i.e., the new image does not have the same orientation as the one used in learning, the required transformation is *T24, which* still needs scaling/perspective calibration based on T_{13} and T_{12}.

Facial Region Transformation

The stored transformations only show the mapping vectors for feature lines. Non-feature points are mapped by interpolating the mapping values for the feature lines surrounding their regions. This is done based on the face region to which a point belongs.

Face regions are grouped into two different categories:

- Feature islands, surrounded by one or two "inner" feature lines
- Face patches, covering the rest of the face as shown in Figure 4-b.

The mapping vector for each point inside a group-1 region is found using the following formula:

$$\mathbf{d}_{r,c} = w(\mathbf{d}_{u,c}, \mathbf{d}_{l,c})$$

where the function *w* is a weighted average with distance as the weights, *r* and *c* are row and column in the image for the given point, *u* and *l* are the row number for top and bottom feature points, and ***d*** is the mapping vector.

Face patches are defined based on covisibility, i.e., their points are most likely to be seen together. Defining the patches is necessary in order to preserve the geometric validity of the transformation. The mapping vector of each point in a patch is the weighted average of mapping of all the patch corners. Extra checking is performed to make sure a point inside a patch will be mapped to another point in a corresponding patch of the target image.

Sample Images

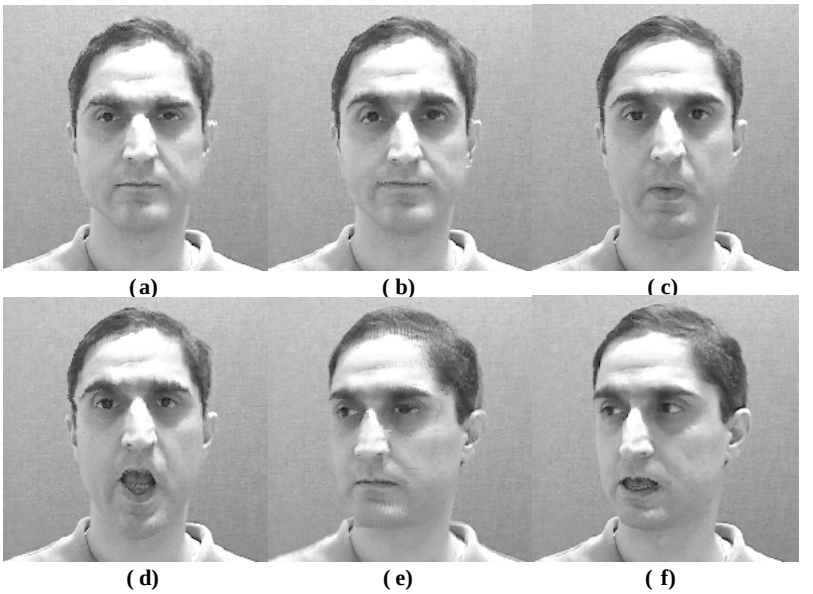
Figure 14 shows some sample outputs of the image transformations.

Speech Synthesis

To achieve the best quality with minimum database requirements, *ShowFace* uses a concatenative approach to speech synthesis. Diphones (the transitions between the steady-state of a phoneme to the steady-state of another) are used as the basis of this approach. An off-line diphone-extraction tool is designed to create a database of diphones from existing audio footage. This database is normalized for power and pitch to provide a smooth initial set. The diphones are then dynamically connected based on the phoneme list of a given text to be spoken.

An FFT-based comparison of diphones finds the best connection point for two diphones at run time. This results in a dynamic time length calculation for diphones and words, which will then be used to find the necessary duration of the corresponding visual transitions and the number of frames to be generated, in order to achieve a lip-synchronized audio-visual stream.

Figure 14. Transformed faces—Mapped from 7-a: (a) frown, (b) smile, (c and d) visemes for sounds "oo" and "a" in "root" and "talk", (e) rotate to right, and (f) non-frontal talking



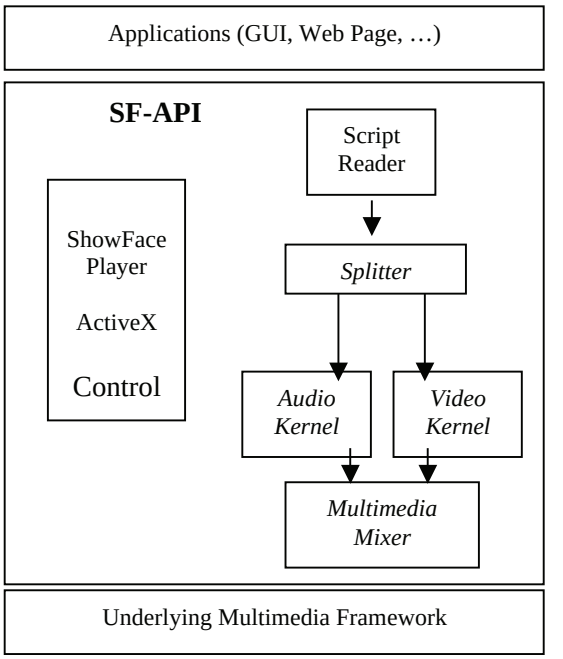
SHOWFACE FRAMEWORK

System Architecture

The basic structure of the *ShowFace* system is illustrated in Figure 15. Five major parts of this system are:

- Script Reader, to receive an FML script from a disk file, an Internet address, or any text stream provider
- Script Parser, to interpret the FML script and create separate intermediate audio and video descriptions (e.g., words and viseme identifiers)
- Video Kernel, to generate the required image frames
- Audio Kernel, to generate the required speech
- Multimedia Mixer, to synchronize audio and video streams

Figure 15. Component-Based ShowFace System Structure



ShowFace relies on the underlying multimedia technology for audio and video display. The system components interact with each other using the ShowFace Application Programming Interface, SF-API, a set of interfaces exposed by the components and utility functions provided by the *ShowFace* run-time environment. User applications can access system components through SF-API, or use a wrapper object called *ShowFacePlayer,* which exposes the main functionality of the system, and hides programming details.

ShowFace system is designed and implemented with the concept of openness in mind. By that we mean the ability to use and connect to existing standard components and the independent upgrade of the system modules. To make most use of existing technologies, *ShowFace* components are implemented as Microsoft DirectShow filters.

DirectShow is a multimedia streaming framework, which allows different media processing to be done by independent objects called *filters,* which can be connected using standard Component Object Model (COM) interfaces. DirectShow will be installed as part of many application programs like browsers and games and comes with a set of filters like audio and video decoders and renderers. This allows *ShowFace* objects to access these filters easily and rely on multimedia streaming services provided by DirectShow, e.g., receiving data from a URL reader or MPEG-4 decoder and sending data to a video player or file writer.

The *ShowFacePlayer* wrapper object is implemented as an ActiveX control, which can be easily used in web pages and other client applications. An off-line tool, *ShowFace Studio,* is also developed to assist in detecting the features, creating the maps and recording the FML scripts. Some samples of transformed faces are shown in Figure 4.

CONCLUDING REMARKS

Reviewing the most important works in the area of multimedia specification and presentation, it's been shown that a comprehensive framework for face animation is a requirement, which has not been met. Such a framework needs to provide:

- a structured content description mechanism,
- an open modular architecture covering aspects from getting input in standard forms to generating audio/video data on demand,
- efficient algorithms for generating multimedia with minimum use of existing footage and computational complexity.

An approach to such a framework, ShowFace System, is proposed. Unlike most of the existing systems, ShowFace is not limited to an off-line application or a media player object, but provides a complete and flexible architecture. The component-based architecture uses standard interfaces to interact internally and also with other objects provided by an underlying platform, making maximum use of existing technologies like MPEG-4, XML, and DirectX. These components can be used separately or in a combination controlled by the animation application.

An XML-based Face Modeling Language (FML) is designed to describe the desired sequence of actions in form of a scenario. FML allows event handling and a sequential or simultaneous combination of supported face states and will be parsed to a set of MPEG-4 compatible face actions. The Main contributions of FML are its hierarchical structure, flexibility for static and dynamic scenarios, and dedication to face animation. Compatibility with MPEG-4 and the use of XML as a base are also among the important features in the language. Future extensions to FML can include more complicated behaviour modeling and better coupling with MPEG-4 streams.

The image-based transformations used in the video kernel are shown to be successful in creating a variety of facial states based on a minimum input of images. Unlike 3-D approaches, these transformations do not need complicated modeling and computations. On the other hand, compared

to usual 2-D approaches, they do not use a huge database of images. They can be extended to include facial textures for better results and the system allows even a complete change of image generation methods (e.g., using a 3-D model), as long as the interfaces are supported.

Better feature detection is a main objective of our future work, since any error in detecting a feature point can directly result in a wrong transformation vector. This effect can be seen in cases like eyebrows where detecting the exact corresponding points between a pair of learning images is not easy. As a result, the learned transformation may include additive random errors, which cause non-smooth eyebrow lines in transformed feature sets and images.

A combination of pre-learned transformations is used to create more complicated facial states. As discussed, due to the perspective nature of head movements, this may not be a linear combination. Methods for shrinking/stretching the mapping vectors as a function of 3-D head rotation are being studied and tested. Another approach can be defining the mapping vectors in terms of relative position to other points rather than numeric values. These relational descriptions may be invariant with respect to rotations.

REFERENCES

Ankeney, J. (1995). Non-linear editing comes of age. *TV Technology*, (May).

Arya, A. & Hamidzadeh, B. (2002). ShowFace MPEG-4 compatible face animation framework. *Int. Conf. Computer Graphics and Image Processing (CGIP)* Hawaii.

Battista, S., et al. (1999). MPEG-4: A multimedia standard for the third millennium. *IEEE Multimedia*, (October).

Blanz, V. & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *ACM SIGGRAPH.*

Bregler, C., et al. (1997). Video rewrite: Driving visual speech with audio. *ACM Computer Graphics.*

Bulterman, D. (2001). SMIL-2. *IEEE Multimedia*, (October).

Cassell, J. et al. (2001). BEAT: The behavior expression animation toolkit. *ACM SIGGRAPH.*

Ekman, P. & Friesen, W.V. (1978). *Facial Action Coding System.* Consulting Psychologists Press Inc.

Ezzat, T. & Poggio, T. (1998). MikeTalk: A talking facial display based on morphing visemes. *IEEE Conference on Computer Animation.*

Funge, J., et al. (1999). Cognitive modeling: Knowledge, reasoning, and planning for intelligent characters. *ACM SIGGRAPH.*

Graf, H.P., et al. (2000). Face analysis for the synthesis of photo-realistic talking heads. *IEEE Conference on Automatic Face and Gesture Recognition.*

Hirzalla, N., et al. (1995). A temporal model for interactive multimedia scenarios. *IEEE Multimedia*, (Fall).

Kallmann, M. & Thalmann, D. (1999). A behavioral interface to simulate agent-object interactions in real time. *IEEE Conference on Computer Animation.*

Lawton, G. (2000). Video streaming. *IEEE Computer*, (July).

Lee, W. S., et al. (1999). MPEG-4 compatible faces from orthogonal photos. *IEEE Conference on Computer Animation.*

Little, T.D.C. (1994). Time-based media representation and delivery. In J.F. Koegel Buford (Ed.), *Multimedia Systems.* ACM Press.

Nack, F. & Lindsay, A.T. (1999). Everything you wanted to know about MPEG-7. *IEEE Multimedia*, (July).

Pandzic, I.S. (2001). A web-based MPEG-4 facial animation system. *International Conference Augmented Virtual Reality & 3D Imaging.*

Pighin, F., et al. (1998). Synthesizing realistic facial expressions from photographs. *ACM SIGGRAPH.*

Tiddeman, B., et al. (2001). Prototyping and transforming facial textures for perception research. *IEEE CG&A*, (September).

This work was previously published in Multimedia Systems and Content-Based Image Retrieval, edited by S. Deb, pp. 356-375, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.12
Multimedia Computing Environment for Telemedical Applications

V. K. Murthy
University of New South Wales, Australia

E.V. Krishnamurthy
Australian National University, Australia

INTRODUCTION

Telemedicine (in short, e-medicine) is a means of delivering medical services to any place, no matter how remote, thereby removing the limitations of space and time that exists in today's health-care settings. Computers are indispensable in telemedicine, since they provide for efficient, relevant data gathering for large-scale applications. Besides providing immediate feedback of results to patients and doctors, they also can compare past patient records and evaluate relative improvement or deterioration. Further, they are readily available at any time, fatigue-free and can be more objective.

Five important application areas of telemedicine are:

1. Lifetime health care;
2. Personalized health information;
3. Tele-consultation;
4. Continuing Medical education; and
5. Context-aware Health monitoring.

For example, computers provide for multimedia imaging: ultra sound, digital X-rays, 3D spiral Cat Scanner, magnetic resonance imaging, PET scanning, and so forth, and can fuse them into a single multi-purpose image using fusion software. Adding mobility to computers enhances their role in telemedical applications considerably, especially at times of emergency since the patients, doctors, the data collecting and retrieval machines, as well as their communication links can always be on the move. Very simple, inexpensive mobile communication and computing

devices can be of great help in telemedicine, as explained in the following:

- ***Low Cost Radio:*** Even the simplest of mobile devices, such as a low power radio that can transmit messages to a home computer from which medical data can be sent through telephone line and the Internet can be of great value in saving lives (Wilson et al., 2000).
- ***Personal Digital Assistants (PDA):*** The simplest of the computers, such as palmtops and PDA can assist the doctors for instant nomadic information sharing, and look for diagnosis of different diseases and treatment. PDA can help the doctors to figure out drug interactions, storing summaries of sick patients and their drug list. Further, PDA can provide for downloading suitable programs from the Web, and can be programmed for alert, sending and receiving email, jotting down pertinent points, and for storing immediately needed clinical results to carry out ward rounds. Also a hand held system can provide context-awareness to support intensive and distributed information management within a hospital setting (Munoz et al., 2003).
- ***Internet:*** The Internet is an important tool for medical professionals and will completely change the manner in which medical consultations are provided (Coiera, 1997); for further details on telehealth and telemedicine practice and their real life implementation issues, refer to Orlov and Grigoriev (2003), Jennett and Anddruchuk (2001), and Suleiman (2001).

For minor ailments, Internet-based consultations to doctors can provide prescriptions for medical/pathological examinations by laboratories. The results are then posted in the Internet for subsequent reading of the results by the concerned doctors who can prescribe medicines that can be posted on the Internet. This prescription can then be handled by a pharmacy to dispense the medicines to the concerned individual. Kim and Hwang (2001) have proposed a password controlled Internet-based medical system that brings in a variety of services to doctors, patients, pharmacists and health-care professionals. It allows people to receive medical examinations and medical advice.

BACKGROUND: TELEMEDICAL INFORMATION SERVICES

The first step in telemedicine is the telemedical diagnosis (or telediagnosis) based on information obtainable from medical images, blood, urine and other pathological test reports. Usually, for diagnostic purposes, the doctor sends a patient for such examinations. The laboratory assistant takes the required X-ray or ultrasound images or carries out pathological tests and passes these images (or readings) on to a radiologist/pathologist who then makes analysis and sends a report to a doctor. These manual actions are totally sequential and slow. This whole procedure can be made cooperative and faster, if the images and data are stored in a database and these can be simultaneously retrieved by specialists in their offices or homes to make a cooperative diagnosis (Alfano, 1997; Coiera, 1997; Ganapathy, 2001; Gomez et al., 1997; Jameson et al., 1996; Kleinholz et al., 1994; Lauterbach et al., 1997).

Principal Aims

The principal aims of e-medical informatics are to:

(i) provide online services of patient records (medical and pathological databases) to medical practitioners and radiologists;
(ii) provide primary specialist diagnosis, offer second opinion, provide pre- and post treatment advice through email;

(iii) reduce the cost of imaging equipment, delay, and increase the speed and volume of diagnosis;
(iv) aid cooperative diagnosis and provide assistance for remote surgery;
(v) provide student /resident education;
(vi) reduce professional isolation, increase collaboration; and
(vii) provide home-care.

Advantages

E-medicine offers the following advantages:

(i) Provides health care to under-served and isolated areas so that we can make a better allocation and utilisation of health resources.
(ii) Since communication cost is much cheaper than the transportation cost, patients in remote areas can outreach physicians quickly.
(iii) Increases the speed of diagnosis and treatment especially when used for teleradiology, cardiology, psychiatry.
(iv) Allows access to specialty care using time-oriented clinical data.
(v) Real-time monitoring of public health databases to prepare and respond during epidemics, biological and chemical terrorism.
(vi) Internet can provide the following support:
    a. Health information;
    b. Administrative infrastructure;
    c. Online health records;
    d. Pharmaceutical information and sales outlets; and
    e. Online training for telemedical professionals.

Prerequisites

The pre-requisites for a successful implementation of a telemedical system are:

- *Infrastructure:* A suitable infrastructure of health care providers, doctors, engineers, computing specialists, communication engineers, information technology professionals and medical statisticians to analyse outcomes and suitable outreach clinics with telemedical facilities.
- *Communication Network:* Reliable, inexpensive and readily accessible communication network from outreach clinics to hospitals, doctors and patients and pathological laboratories.
- *Low-cost Computers:* Suitable low-cost hardware-software and a good communication bandwidth to transmit medical data in different modes; for example, radiological images, video images of signals and text. While using wired in or wireless mobile devices and monitors, the effect of electromagnetic interference (EMI) and radio frequency interference (RFI) on data collection and transmission, and the side-effects on patients, both physiological and psychological aspects, have to be taken care of so that improper diagnosis does not result.
- *Training Facility:* Training of personnel to provide proper maintenance of equipment and safety standards to patients.
- *Security, Reliability, Efficiency:* Reliability, Efficiency, Security, Privacy and Confidentiality in handling, storing and communicating patient information.

Economic Necessity

In densely-populated countries (e.g., India), the rate of growth in hospital beds to cope up with the increasing population is economically unsustain-

able and technically not viable, since the number of medical specialists also cannot grow to meet this demand (Ganapathy, 2001). The use of telemedicine avoids unnecessary strain involved in travel and associated expenses, provide immediate attention and care, can avoid hospitalization, and allow the patients to stay home enjoying family support. Chan et al. (2000) describe a real-time tertiary foetal ultrasound telemedical consultation system using standard integrated system digital network (ISDN) that operates in Queensland, Australia. This consultation system has gained acceptance from the clinicians and patients. Aging population and rising health costs have created the need to care for more patients in their own homes. Hospitals without walls (e-hospitals or virtual hospitals) provide for continuous monitoring of patients in certain diagnostic categories. Wilson et al. (2000) describe how to build such "hospital without walls". This system uses a miniature, wearable low power radio to transmit vital and activity information to a home computer, from which data is sent by telephone line and the Internet to the concerned doctors. Thus, telemedicine and telediagnosis are economic necessities for both the developing and the developed world.

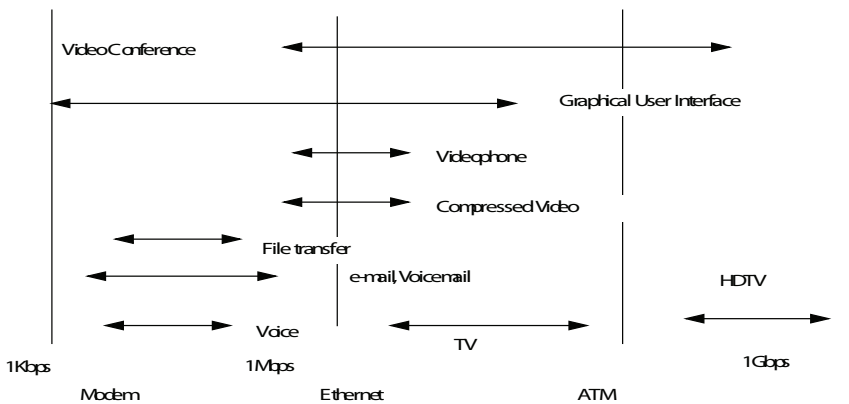
MAIN THRUST OF THE ARTICLE

The main thrust of this chapter is to describe the multimedia telediagnostic environment (MMTE), its features, and some practical MMTE systems that are currently operational.

MULTIMEDIA TELEDIAGNOSTIC ENVIRONMENT

A Mobile Multimedia Telediagnostic Environment (MMTE) permits users to work remotely on common shared resources and applications and simultaneously communicate both visually and through audio. Such an environment becomes more flexible and useful if it can work in a wireless and wired-in (integrated) environment so that the services can be provided for mobile hosts. Here, the users have interfaces of PDA or laptops interacting remotely. In this application, we need cooperation among the participants through special communication tools for conversation (Roth, 2001) and E-pointing (MacKenzie & Jusoh, 2001). The conversation can be of conference type where two or more participants are involved. The e-pointers aid each participant to point out a particular segment of an image or a

Figure 1. Bandwidth (log scale) requirements



video image of a signal so that each of the other participants can visualize the movements of the remote pointer. Also in telesurgery (e-surgery) where two or more operating rooms are connected by a network, live video signals may required to be transferred to other remote locations from endo-camera (camera for photographing internal organs) and operating room camera to other remote locations for consultancy. This would allow surgeons not only to see but also visualize surgical instrument movements with 3-D models in real time during the surgery.

Pre-requisites for the Deployment of MMTE

Reliable Communication Network and Equipment

A reliable multimedia communication network that links the remote centres with hospitals is essential. The delivery of multimedia content in a timely fashion is very important. Internet provides a uniform communication medium to link users together to access or deliver multimedia information.

- Communication Bandwidth Requirements: Text requires the lowest bandwidth, while audio and video data and signals require significant increase in bandwidth. Specification of bandwidth to be used and compression techniques used are to be laid down so that the data and images that are transmitted are of diagnostic quality.
- Bandwidth Management: Bandwidth determines the information capacity of a network per unit of time. Wireless networks deliver lower bandwidth than wired network. Hence software techniques based on compression should be used. Also scheduling communication intelligently can save bandwidth. For use in telemedicine the techniques should be extremely reliable. Current cutting edge technologies are yet to develop. Bandwidth requirements along with applications are given in (approximate) logarithmic scale in Figure 1.

Three common technologies used are: (bps = bits per second; K= Kilo; M= Mega; G = Giga)

Dial up mode: Rate 28.8Kbps
T1: 1.544 Mbps; DS3: 45 Mbps

For example, standard X-ray transmission takes 30 minutes in dial-up mode, 30 seconds in T1 mode and 1 second in DS3. It is obvious that the cost goes up, as we want to increase the bandwidth to communicate voice, data and pictures. DS3 is a dedicated, private line service designed for point to point communication. This service uses fibre optic cable. One can have the option of tiered DS3 service from lower to higher bandwidth from 5 Mbps to 45 Mbps depending upon cost considerations and transmission requirements.

- Choice of Multimedia Hardware: We need a proper choice of graphics and audio equipment for quality images and audio for diagnosis.

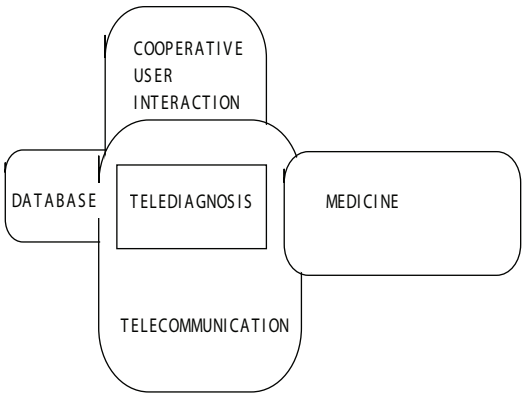
FEATURES OF A TELEDIAGNOSTIC ENVIRONMENT

A typical cooperative telediagnostic (e-diagnostic) environment is shown in Figure 2. It consists of databases containing medical/pathological image and textual information, cooperative interaction facility, and telecommunication. Typically, its characteristics are:

Remoteness of Users

Users work remotely on common shared resources and applications and simultaneously communicate both visually and through audio. This environment becomes more flexible and useful, if it can

Figure 2. Cooperative diagnostic mobile environment



work in an integrated (wireless and wired-in) environment so that the services can be provided for mobile hosts.

Cooperation

Many e-medicine applications require doctors, specialists and health-record officials to cooperate and discuss for particular medical treatment. Cooperation among the participants requires special communication tools for conversation, e-pointing, e-surgery.

E-pointing Facility

E-pointing is the act of moving an on-screen tracking symbol such as a cursor, by manipulating the input device. The interactive e-pointers (MacKenzie & Jusoh, 2001) aid each participant to point out a particular segment of an image or a video image of a signal so that each of the other participants can visualize the movements of the remote pointer and discuss any abnormality for diagnostic purposes.

Facility for Internet Link

Also users can remotely access worldwide knowledge bases/databases, download files from the Internet and browse the World Wide Web using portable and mobile devices such as laptops, palmtops and PDA.

Interoperability

Software interoperability is a key requirement in telemedicine, since different modalities of patient records are used on different platforms.

- **Security Requirements**: Security of patient data during storage and transmission are vital to safeguard confidentiality and privacy of patient information. Biometric authentication (Nanavati et al., 2002) could be an answer in dealing with patient information. Health Insurance Portability and Accountability Act (HIPAA) has devised certain security measures to handle telemedical information, see Web document of HIPAA.

FUTURE TRENDS: SOME PRACTICAL MMTE SYSTEMS

Context-Aware Mobile Communicaton in Hospitals

Handheld systems can provide for context-awareness in hospitals. This can provide speedy collaboration among doctors to deal with time sensitive cases, for example, delivery, heart attacks, accidents (Munoz et al., 2003).

Presently, portable and mobile devices such as laptops and personal digital assistants (PDA) are useful for remotely accessing resources in an efficient and reliable manner and for reliable telediagnosis (Lauterbach et al., 1997; Roth, 2001). For instance, information sharing among doctors through handheld appliances are useful for transmitting small amounts of data, such as heart rate, blood pressure and other simple monitoring devices. Typical datatypes for current handheld

appliances are: text, data entries, numbers, tables. None of the applications described previously can deal with multimedia data, such as audio, video which require considerable bandwidth, sufficient output devices, and a very powerful battery. Presently, multimedia data are not yet suitable for handheld appliances; they need to undergo major developments (Roth, 2001). Currently available tools, such as GyroPoint and RemotePoint (MacKenzie & Jusoh, 2001), lack accuracy and speed. The telepointer technology has yet to mature to provide reliable service for e-medicine and e-surgery. Along with the use of virtual reality, the pointer technology can be useful in teaching surgery, and for trial-error planning of operations.

Image-Guided Telemedical System

Lee (2001) describes a Java-applet based image-guided telemedicine system via the Internet for visual diagnosis of breast cancer. This system automatically indexes objects based on shape and groups them into a set of clusters. Any doctor with Internet access can use this system to analyze an image and can query about its features and obtain a satisfactory performance (http://dollar.biz.uiowa.edu/~kelly/telemedicine.html).

Software Agent based Telemedical Systems

Software agents, which are personalized, continuously running and semi-autonomous objects, can play an important role in cooperative telemedicine. Such agents can be programmed for supporting medical diagnostic intelligence and keep a watchful eye to discover patterns and react to pattern changes that usually occur in epidemics, biological and chemical terrorism.

For example, diabetes is a chronic disease with a sustained elevated glucose level. The diabetics need to be injected insulin to control the glucose level. This is quite a tedious book-keeping procedure. Real-time monitoring using agents and immediate therapy can save a life.

Also agents can help in the appropriate choice of doctors, allocation of hospitals and beds for patients. Ganguly and Ray (2000) describe a methodology for the development of software agent-based interoperable telemedicine system. This system has been illustrated using tele-electrocardiography application. Also recently, a practical agent-based approach to telemedicine has been suggested by Tian and Tianfield (2003). For medical education, this will be a very useful tool. The Foundation for Intelligent Physical Agent Architecture (FIPA, http://www.fipa.org) aims to improve agent interoperability by providing standards for protocols and languages. Also the Java 2 MicroEdition (J2ME) is targeted at PDA. These developments will provide an agent execution environment in PDA.

CONCLUSION

Cooperative telemedical informatics will have a direct impact on the rapidly changing scene of information and communication technology (including advanced communication systems) and will provide for greater interaction among doctors and radiologists and health-care professionals to integrate information quickly, efficiently and make effective decisions at all levels. Virtual e-hospital or hospital without walls will be a reality soon providing great benefit to the society.

REFERENCES

Alfano, M. (1997). User requirements and resource control for cooperative multimedia applications, *Multimedia Applications, Services and Techniques, Lecture Notes in Computer Science*, 1242, 537-553. New York: Springer Verlag.

Chan, F.Y. et al. (2000). Clinical value of real-time tertiary foetal ultrasound consultation by telemedicine. *Telemedicine Journal, 6*, 237-242.

Coiera, E. (1997). *Guide to medical informatics, the Internet and telemedicine*. London: Chapman & Hall.

Ganapathy, K. (2001).Telemedicine in action •The Apollo experience. *Proceedings of the International Conference on Medical Informatics*, Hyderabad, India, November 2001.

Ganguly, P. & Ray, P. (2000). A methodology for the development of software agent based interoperable systems: A tele-electrocardiography perspective. *Telemedicine Journal, 6*, 283-294.

Gomez, E.J. et al. (1997). The Bonaparte telemedicine. In *Multimedia Applications, Services and Techniques, Lecture Notes in Computer Science*, 1242, 693-708. New York: Springer Verlag.

Jameson, D.G. et al. (1996). Broad band telemedicine teaching on the information superhighway. *Journal of Telemedicine and Telecare, 1*, 111-116

Jennett, A., & Anddruchuk, K. (2001). Telehealth: Real life implementation issues. *Computer Methods and Programs in Biomedicine, 64*,169-174.

Kim, S.S. & Hwang, D.-J. (2001). An algorithm for formation and confirmation of password for paid members on the Internet-based telemedicine. *Lecture Notes in Computer Science*, 2105, 333-340. New York: Springer Verlag.

Kleinholz, L. et al. (1994). Supporting cooperative medicine. *IEEE Multimedia*,Winter, 44-53.

Lauterbach, Th. et al. (1997). Using DAB and GSM to provide interactive multimedia services to portable and mobile terminals. In *Multimedia Applications, Services and Techniques, Lecture Notes in Computer Science*,1242, 593-609. New York: Springer Verlag.

Lee, K.-M. (2001). Image-guided telemedicine system via the Internet. *Lecture Notes in Computer Science*, 2105, 323-333. New York: Springer Verlag.

MacKenzie, S. & Jusoh, S. (2001). An evaluation of two input devices for remote sensing. In *Engineering for Human Computer Interaction, Lecture Notes in Computer Science*, 2254, 235-250. New York: Springer Verlag.

Munoz, M.A. et al. (2003). Context-aware mobile communication in hospitals. *IEEE Computer, 36*(9), 38-47.

Nanavati, S. et al. (2002). *Biometrics*. New York: John Wiley.

Orlov, O., & Grigoriev, A.(2003). Space technologies in routine telemedicine practice: Commercial approach. *Acta Astronautia, 51*(July), 295-300.

Roth, J. (2001). Information sharing with hand-held appliance. In *Engineering for Human Computer Interaction, Lecture Notes in Computer Science*, 2254, 263-279. New York: Springer Verlag.

Suleiman, A.B. (2001). The untapped potential of telehealth. *International Journal of Medical Informatics*, 61, 103-112.

Tian, J. & Tianfield, H. (2003). A multi-agent approach to the design of an e-medicine system. *Lecture Notes on Artificial Intelligence*, 2831, 85-94. New York: Springer Verlag.

Wilson, L.S. et al. (2000). Building hospitals without walls: A CSIRO home telecare initiative. *Telemedicine Journal, 6*, 275-281.

WEB-DOCUMENTS

http://telemed.medicine.uiowa.edu: Zollo Susan: Introduction to Telemedicine

http://www.nlm.nih.gov/reserach/telesymp.html: Provides information on Telemedicine Symposium.

http:// telemed.medicine.uiowa.edu: Provides slide shows on various aspects of telemedicine prepared by the National Laboratory for the study of Rural telemedicine, of the University of Iowa.

http://telemed.medicine.uiowa.edu/M.G.Kienzle, Telemedicine for home monitoring

http://www.ntia.doc.gov/reports/telemed/, Telemedicine report to Congress.

http://telemed.medicine.uiowa.edu/ (UI TRC).

http://tie.telemed.org/(TelemedicineInformation Exchange).

http://www.nlm.nih.gov/research/telemedinit.html (NLM National Telemedicine Initiative).

http://www.tmgateway.org/ (Federal Telemedicine Gateway

http: //www. hipaadivsory.comregs/securityoverview.htm

KEY TERMS

Bandwidth Management: Determines the information capacity of a network per unit of time. Wireless networks deliver lower bandwidth than wired network. The choice of appropriate bandwidth for efficient and cost effective transmission of voice, data and pictures is called bandwidth management.

Confidentiality and Security in Telemedicine: Security and confidentiality of patient data during storage and transmission are vital to safeguard confidentiality and privacy of patient information. Biometric authentication, could be an answer in dealing with patient information. HIPAA (Health Insurance Portability and Accountability Act) has devised certain security measures to handle telemedical information (http://www.hipaadivsory.comregs/securityoverview.htm)

Multimedia Telediagnostic Environment (MMTE): Permits users to work remotely on common shared resources and applications and simultaneously communicate both visually and through audio.

Software Agents: Personalized, continuously running and semi-autonomous objects. They can play important role in cooperative telemedicine. Such agents can be programmed for supporting medical diagnostic intelligence and keep a watchful eye to discover patterns and react to pattern changes that usually occur in epidemics, biological and chemical terrorism

Software Interoperability: Permits different software to run on different platforms with different operating systems.

Telemedicine (E-Medicine): A means of delivering medical services to any place, no matter how remote, thereby removing the limitations of space and time that exists in today's health-care settings.

Telepointers: The act of moving an on-screen tracking symbol, such as a cursor, by manipulating the input device from a remote site.

Virtual E-Hospital (or Hospital Without Walls): A telemedical facility that provides continuous monitoring of patients staying in their homes and enjoying family support and creates a psychological satisfaction to the patients that they are receiving immediate attention and care as if they are in a real hospital.

This work was previously published in Mobile Commerce Applications, edited by N. S. Shi, pp. 95-116, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.13
Interactive Multimedia and AIDS Prevention:
A Case Study

José L. Rodríguez Illera
University of Barcelona, Spain

ABSTRACT

Using multimedia applications to inform or to train is very different than using them for changing attitudes. The documented and discussed project started with the perspective that a large proportion of young people, despite knowing how AIDS might be contracted, still adopt risk behaviors. A multimedia role play application was designed to include both information and game layers. The game introduces complex situations using video stories, and then lets the users construct different narratives by choosing between behavior alternatives. The result of each narrative is related to contracting the disease or not. A discussion about role playing games follows, on the limits of this approach, as well as the kind of interactivity and the forms of delayed feedback given.

INTRODUCTION

This chapter provides a detailed description of a multimedia AIDS prevention project undertaken jointly by research teams in Italy and Spain. The project, "AIDS: Interactive Situations," was funded by the European Union and resulted in the setting up of a Web site and the production of a hybrid CD-ROM, of which more than 40,000 copies were distributed, through both public and private channels, in the two participating countries between 1999 and 2000. The chapter is divided in five parts: a description of the project rationale and an outlining of its objectives; a description of the project's contents; a description of the multimedia technology used and the interactive approach incorporated; a discussion of the project; and conclusions reached.

PROJECT RATIONALE AND OBJECTIVES

AIDS prevention is a constant concern of the education and health authorities. Prevention campaigns are frequently mounted, and wide use of the mass media is made in conveying the message. However, interactive media have only rarely been used for this purpose.

At the start of the 1990s, the only software available were HyperCard stacks and similar programs containing AIDS fact files and information about the ways in which the disease might be contracted, and a number of simulation programs based on system dynamics models that demonstrated the evolution of the disease at a time when it was thought to be fatal in a period between 10 and 15 years (González, 1995). Multimedia programs were later developed, but their primary purpose was as a source of medical information (AIDS 2000 Foundation). Other programs included a computer game that allowed the study of epidemics throughout history (Fundació LaCaixa, 1995).

In developing this project, "AIDS: Interactive Situations," the aim was to provide a different focus. In fact, by the mid-1990s, most adolescents (here, and throughout the chapter we refer solely to adolescents in the Western world) had a good grounding in the basics of AIDS prevention, thanks in large measure to the prevention campaigns. Yet, despite knowing how the disease might be contracted, a large proportion of adolescents still adopted risk behaviors. This discrepancy between the information received and the attitudes that guide their behavior is a constant feature among adolescents.

The main aim of this project was, therefore, to focus on the subjects' perceptions of risk situations and the consequences of their behaviors. The other objective of the project involved the provision of decision-making techniques in situations of risk, always exemplified by the failure to use a condom in heterosexual relations.

CONTENTS AND EDUCATIONAL DESIGN

The results of the project's psychological and educational analyses indicated the type of contents and transformations required. We concluded that the best approach was to include a purely informative content, offering information about the disease and the ways in which it might be transmitted, plus information describing its psychological and social features. This information serves as a ready reference for schools and can also be consulted on an individual basis. As we shall see later, it serves an additional function, one that we consider to be of considerable importance. This information "layer" is included in a straightforward hypertext format and aims above all to be user friendly. It also incorporates a number of further multimedia tools, including a map of AIDS information centers. The contents are organized in five sections:

1. *The disease*: This section contains information about HIV, how the virus is produced, how it acts on the organism, etc.
2. *Prevention*: This section allows the user to acquire information about how to prevent AIDS, which sexual practices involve the greatest and least risk, and how to use male and female condoms.
3. *The AIDS test*: This section explains when one should take the AIDS test, how to go about taking it, and how to interpret the results.
4. *Ideas and behaviors*: This section focuses on techniques to become more assertive, including negotiation and dialogue at times of conflict, understanding oneself better, etc.
5. *To find out more*: This section lists books, songs, films, and Internet Web sites that contain information about AIDS.

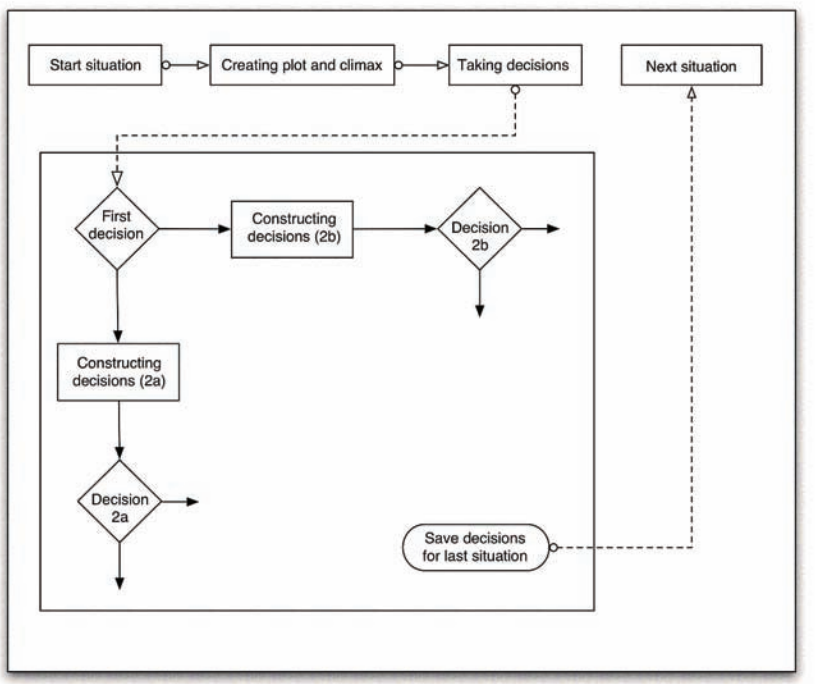
The main content, however, comprises an interactive role-playing game. This format was selected as it was considered the best way to meet the aims of changing attitudes and of simulating the negotiation and dialogue that occurs in situations of risk. In Tonks' (1996) review of techniques for providing information about AIDS and changing attitudes about the disease, role-playing games

appear as the basic tool, although not as part of a multimedia application—which in Tonk's review are considered only in their audiovisual format. Role-playing games offer many advantages, above all the possibility of testing the skills that are being learned or developed in a safe environment. Furthermore, role play allows great flexibility in terms of content, and it is typically used without any multimedia components.

The role play is based around the metaphor of a summer trip taken by a group of friends around Europe. In this group, there is a couple in a steady relationship who have to deal with a number of different situations. The program user has to choose at the outset whether to be the male or female character and has to behave in accordance with this choice throughout the journey, as the content varies depending on the role that has been selected. The choice of character does not depend on the sex of the player, given that the game can be played in a group or as part of a class activity within a school, but it conditions the way in which the game develops, presenting a particular point of view in each situation. In fact, we believe that this initial choice constitutes the user's main point of identification with the game, because the player then has to interact in the program as if he or she were one of the characters and to adopt what they consider to be the character's point of view.

The role play is organized around six situations: the first acts as an introduction, the next four present risk situations, and the last tells the user the results of the decisions he or she has taken. Each of the four risk situations is organized in a similar manner: first, a narrative section is presented in which a complex situation is introduced, followed by an interactive section in which decisions are made. This common format ensures that the main story line in the game is easily followed, as the metaphoric journey is always brought to a halt by a situation that is presented in a similar way, and after the user has made the required decisions, he or she can continue on the journey, whatever happens. Decisions have to be taken: the user

Figure 1. Flow diagram of the game



cannot proceed with the game if decisions are not made, and the user's results are stored and not shown until the end of the game. Figure 1 shows the overall organization of the game, although the decision-making tree diagram only shows the first two levels.

Each situation involves the use of condoms in heterosexual relationships, but each emphasizes a different ability that we wish to strengthen within the general framework of negotiating condom use: the first is the ability to stand by one's opinions in a dialogue with one's partner, the second is resisting group pressure, the third includes the situations that arise when changing partners, and the fourth includes the decisions that are made under the influence of alcohol and drugs. In addition to these main abilities, each situation presents a considerable amount of informative material contained within the dialogues, both in the video and in the decision-making section. This information, at times debatable as it is presented as the opinion of one of the characters, is contained within the project's hypertext.

The choice of content is important in several respects: first because of the abilities described above, second because of the physical settings in which the story unfolds, third because of the overall credibility of the situations, and fourth because of the language presented within the dialogues.

The Educational Design

Given the complexity of the psychological and pedagogical aspects of the project, it is extremely difficult to find one conceptual framework that can support its educational design. In general, educational multimedia applications tend to use a cognitive theoretical framework or, on occasion, a constructivist one (Duffy et al., 1992; Duffy & Cunningham, 1996). However, in most cases, applications are not designed on the basis of a single theoretical viewpoint but use several to try to resolve a specific instructional problem.

The main feature that distinguishes multimedia projects such as the present one from approaches that seek to automate instructional design is that they are driven both by the problem and by the theoretical frameworks of the designers. That is, inside specific theoretical orientations, instructional and learning strategies are sought that make it possible to resolve the problem—a *bricolage*-type activity. Determining what is most important is only possible if the characteristics of each particular case are taken into account.

For the most part, multimedia role plays such as "AIDS: Interactive Situations" use general educational principles of constructivist type, albeit in combination (*bricolage*) with other related approaches. These principles have been studied on many occasions, though the analysis has been largely generic and has not been applied to specific cases of educational multimedia applications. We can speak of three fundamental principles that guide the educational design:

1. The individual construction of meaning
2. The situated character of cognition and learning
3. The play environment as a construction of the player's identity

The Individual Construction of Meaning

This is the fundamental principle of constructivist approaches and is what distinguishes them from teaching models that are based on the transmission of knowledge. Knowledge is constructed by integrating meaning (or sense) into preexisting, personal structures.

In our case, the content is structured according to the model of interaction discussed above, that is, by allowing the pupils to select their own paths by means of the choices that they make. Participants construct their own path, or narrative, by choosing from the many alternatives that the game allows. In contrast, many of the programs that merely provide information on AIDS, without

offering other activities (interesting as they are in their own right) can be considered as merely transmitting a particular type of knowledge (medical, psychological, or social). The construction of meaning requires the involvement of the learner so that the new knowledge is integrated and internalized, even in the case of a simple activity such as deciding how a story is going to develop.

The Situated Character of Cognition and Learning

The concept of "situated learning and cognition" (Lave, 1988, 1990) which is a radical critique of the cognitivist vision, stresses the need to place learners in situations that are meaningful to them. It considers that all learning is linked to the social situation or context in which it is produced. In the case under analysis here, this view of learning has developed into the current notion of "learning communities" and finds its expression in the attempt to make the role play a "situated activity." "Situated activity" is an activity that is both meaningful and credible: meaningful because it focuses on a problem that is important to the subjects and credible because it is life-like (in spite of the inevitable fact that it is presented by means of a computer screen).

Credibility is the main characteristic in each of the situations contained within a role play, because it is impossible to get a subject involved if he or she does not consider the situation to be realistic. This realism is achieved by the careful study and use of three types of factors: physical, linguistic, and narrative:

1. ***Physical.*** The situations are credible as far as the physical setting and the characters' ways of dressing and moving are concerned. The actors chosen were of ages between 16 and 20 and were encouraged to give their performances as spontaneously as possible. Furthermore, the interactive situations were played out in settings not unfamiliar to young people: a hotel room, a beach after sunset, a discotheque, and a party in the house of friends.
2. ***Linguistic.*** As in other simulations or pseudosimulations, in role plays such as this, it is the content that conditions the simulation and the realism of the situations, and this is primarily language-based: the game is largely concerned with taking decisions but also with following the reasoning that leads to a decision and, finally, opting between two alternatives that represent opposite, or markedly different, points of view. For this reason, the characters' ways of speaking had to be selected with particular care in order to capture as closely as possible the way young people express themselves.
3. ***Narrative.*** The game's story line is organized around the metaphor of a journey. To ensure realism, the journey involves a group of friends visiting various European cities one summer by train. At each stop, a new situation can be introduced, and in this way, the journey serves as a narrative thread linking each situation (a thread that would have been difficult to find if the situations had occurred as isolated incidents). Having said this, however, each situation is independent of those that precede it and stands as a separate situation in its own right, with its own problem and solution.

The outcomes of each decision are not revealed, however, until some months after the holiday. On the one hand, it needs to be like this in order to give greater realism to the game, as there exists the so-called "window period" during which infection with the AIDS virus cannot be diagnosed, even when it has occurred. On the other hand, it captures the particular characteristic of the apparent disconnection between the risk behavior and the onset of awareness: when the antibody test can be performed, the subjects have forgotten the practices that have led to the results they

are given. In this case, the game indicates the situations and practices of risk during which the infection could have occurred.

The Play Environment as a Construction of the Player's Identity

This aspect is common to games in general, and is particularly relevant in computer games, including role plays. The setting provides a safe environment in which players can experiment with activities that involve a certain risk; they can break the rules in some way, or they can improvise their reactions to an unexpected situation. In role plays, participants represent different personalities and act accordingly but are not liable to suffer from any negative consequences of the decisions they take. The role-play scenario is a safe environment, but it is also a learning environment in which the participants' identities are modified by the ways in which they play the roles of the imaginary characters. This connection between learning and identity has been highlighted by Wenger (1998) and more recently by Gee (2003) with respect to video games.

In the case of AIDS prevention, role play allows participants to create a situation in which they play the parts of adolescents through their identification with their roles, but at the same time without the risk of suffering the negative consequences of the decisions they make. The play environment, the identification with the character, the active choices made in selecting a narrative and constructing meaning, and the "realistic," credible nature of the situations act synergistically in the educational design.

INTERACTIVE MULTIMEDIA APPLICATIONS

The decisions taken regarding project aims and content have a direct bearing on several aspects of the multimedia production, as well as the interactive applications.

Multimedia Production

The multimedia production typically includes the graphic interface design, the media, and the programming. The graphic interface was designed following criteria similar to those adopted in the content specifications, which seek to make them suitable for a young end user. The overall design comprises several points of focus that vary as the journey takes its course, and this, in part, reflects the distinctive sections of the project: the hypertext reflects a more conventional presentation of the project's contents, the role play uses a black background combined with a number of innovations (which are described below in the description of the interactive applications), and, interestingly, a small game serves to introduce the various characters, each using a variation in the graphic interface.

The overriding idea in determining the graphic interface was to come up with a product that was as close as possible to the aesthetic design that young people are used to seeing in computer games, multimedia leisure activities, and even in video clips and television. Unlike many adult users, children and young people are particularly critical of graphic interface features, and many educational programs fail to make any attempt to capture the graphic aesthetics that appeal to them. It is also true that the cultural similarity between the two countries in which the CD-ROM was distributed, Italy and Spain, helped unify the graphic criteria used.

The media used included video (nine sequences of around 5 minutes each, bearing in mind that the last sequence had to be divided in four: two depending on the selection of the sex of the character, one showing infection with the virus, and the other showing a situation in which the disease is not contracted), several hundred stills for the simulated dialogues, and music.

The decision to use video, as well as that to use photographs, was made to promote user identification with the characters in the role

play. Unlike animated images, which have to be extremely realistic or of high quality, video encourages identification with the characters, and with the story, more easily and more directly. The performances of the actors, and their facial expressions, ensure that in the mind of the user the actors and the characters are inseparable. The stills were taken using conventional photographic techniques, predominantly in close up, with some shots taken at a slightly longer range. The reasons for this are well known in the cinema, as close-ups of the face, capturing the actor's facial expressions and eyes, help the viewer identify with the character.

In short, the choice of the media was based on the need to make the story as realistic as possible, and as such, both video and photography were seen as essential elements in capturing the emotional impact of the story.

The computer technology and programs used were conventional: we used Macromedia Director for the design, given its versatility and the ease with which different media can be integrated, in addition to its multiplatform capacity, linked to QuickTime. A special Internet version was not designed, given the video size (an average of 50 MB), which would have meant it could not even be used on wideband networks. Furthermore, most of the users are secondary schools, or citizen support groups, or young people in general, who typically only have access to an ISDN or ADSL modem connection with a capacity to download video images that is extremely limited.

Interaction

At its most basic, the interaction element is organized around a simple navigational structure in which the user must choose between the information section and the role play. Below we shall see how these two options are interconnected. The information section consists of a hypertext comprising graphics and text, which provide basic information. The text is adapted to the users and is extremely user friendly.

The role play, however, has a more complex interactive format, as it combines a story told in video images with the need to make decisions (stills). The video story is interrupted when a conflict arises between the characters, and the user is left not knowing how it will evolve. The audiovisual story serves, then, to motivate the user and also to present an unresolved problem. The user then needs to respond to this problem according to the role he or she has adopted in the game, which is the character with which the user by now identifies. Therefore, the interaction with the content of the program centers on the choice of various options in a simulated conversation with the main characters—depending on which of two options given is selected, the subsequent options presented will vary.

As described above, once the interaction has been initiated, an internal narrative is constructed in accordance with the options selected: the course taken by the dialogue is determined by the choices that are made. In other words, the application itself constructs the narrative and the course taken by subsequent choices, using a preprogrammed dialogue that is inserted between nodes in the decision tree. This dialogue takes the form of various screens, very much in the style of a photo-novel.

One of the most interesting aspects of this system is that it allows the decisions to be thought through. In other words, there is no time pressure whatsoever on the user, who is free to make his or her decision when they feel fit. This means that the dialogues between one decision and the next can be read and given due thought, as they result in the need for a new choice to be made. The application includes the possibility of returning to the video sequence at any time, as well as changing the choices made, should the user feel he or she made a mistake or wishes to select the other option. The decisions taken are depicted in the form of mini graphic images, so that it is always possible to go back to one of them (although, of course, changing the earliest decisions means that all subsequent decisions are lost).

Image 1. Decision-taking structure in the role-playing game. The video still on the left allows the player to view the whole situation. The small circular images show the previous decisions that have been taken. The player can return to these if he or she wishes to reconsider the decision. The two main images in the middle show the options that the player has to choose between for the situation just viewed: by placing the mouse over each image a text appears summarising the option, while the other image fades out. The small image of the main character at the bottom of the screen reminds the player that they have adopted the role of the female character.

Interaction during the game enables the more complex decisions, or those that require factual information, to be linked up with the hypertext system in the information section of the program. If the young user wishes to receive information before taking a decision, he or she can launch the information system, although the hypertext capacities of the system are restricted: it is only possible to navigate those screens containing relevant information for the decision that has to be taken at that moment. This is a design choice, implemented so that the user does not cast the net too wide when searching for information and so as to give contextualized support only.

Educational Applications and User Tests

The project was distributed with the national newspapers and was also sent to educational resource centers. This distribution plan ensured a wide audience but made it difficult to conduct any evaluation of its impact. However, an informal method of evaluation was employed by conducting interviews with the users. The results of this (for a detailed account see Rodríguez Illera et al., 1999) revealed a very high approval rating, while respondents claimed that they had identified easily with the role play. The only criticisms received concerned features of the interface, in particular, in the information section, where some users felt the text was too dense and the letter size too small.

This somewhat limited analysis concerned the programs used in groups with a teacher. For such purposes, the program is accompanied by a detailed guide for educational contexts, one for teachers and another for the end users [http://www.noaids.org]. The latter suggests various activities and means of comprehension for users working by themselves. This possibility was specifically included so as to allow those young people who feel uneasy or who are reluctant to express their opinions in public use the program.

DISCUSSION

We believe that the project demonstrates the means of integrating multimedia capabilities within an instructional design that has clear educational objectives, incorporating elements of interaction to reinforce these objectives within an overall framework that comprises a role-playing game. We would highlight the following aspects of the project:

1. The search for design simplicity at all times: Rather than use multimedia capabilities for their own sake—including animation, audio elements, and music—with little bearing on the educational purpose of the project, we sought to use only those elements necessary to meet the project's specific educational aims. This does not mean that we ruled out the use of more complex interactive capabilities, in particular, given the type of end user we are dealing with. Indeed, the program incorporates a section in which a wide range of multimedia capabilities is used with the primary purpose of entertaining the user: before embarking on the journey, the program allows the user to get to know the main characters of the story better by using a number of short interactive games that differ for each of the six characters. However, this part is clearly isolated from the rest of the program and does not interfere with either the information section or the role-playing game itself.
2. The use of multimedia is designed to facilitate the telling of the story, to create dramatic tension and climatic situations, and to introduce the conflicts. In other words, the features of audiovisual language are used—in this case, features that are more emotive than informative and features that ensure the user identifies with the story's characters.
3. It is true that we do not have an institutionalized means of representing the language of multimedia (Plowman, 1994), and that, therefore, it is difficult to know the significance of certain multimodal configurations (Kress, 2003), such as those that are present in designing complex screens. However, in the case we are concerned with here, the central place of the video in the construction of the story, as well as the absence of the simultaneous appearance of text, means that it can be considered as the dominant component of multimedia, and it can be thought of as being largely responsible for constructing the meaning.
4. The program combines a story, which has its own predefined meaning, with elements of interaction that provide a new meaning and a situation that each user constructs via the decisions that he or she makes, creating a personal narrative along the path that is taken. This format employs interactive multimedia capabilities, while putting them at the service of educational objectives.

The Limits of Role Play

This project description has highlighted what we consider to be its successes, but a subsequent analysis enabled us to see where its limitations lay, in particular, those concerning its instructional design. As indicated, role play is a type of simula-

tion, albeit without any underlying mathematical model, in which it is easy to rehearse certain skills in a safe environment. The strength of the simulation lies in user identification with the characters (and with all the other aspects that are constructed using the multimedia). Once this has been achieved, it becomes virtually independent of the multimedia format that the subsequent interactive program adopts, though not of the logic underlying the choices that are made throughout the program and the story that develops. We believe that this principal feature of the project can be seen in terms of a theoretical construct similar to that present in situated learning and cognition: the attempt to get the young users to see it as something that they have to resolve in a particular way, using the elements that appear on the screen. The "magic" of multimedia applications is in the complete engagement of the user, in a similar way to that in which a book absorbs its reader (Hill, 1999), in other words, its ability to transport the reader or the user to a very high level of cognitive involvement, centered on the activities that they have to carry out. In short, it is the ability to make the user believe that the role play is a real and engaging situation.

The underlying logic of a multimedia role play differs from that of the goal-based scenarios proposed by Schank (1998), which might be considered as another strategy of situated learning. It responds more closely to those cases of ill-defined problems that are so typical of informal teaching and learning situations. The inadequate definition of the situation or the problem is characteristic of real problems, which, removed from experimental situations, need to be analyzed using multiple perspectives and argumentations and descriptions designed to capture their meanings. In the case of role plays, this need is apparent in the narratives of decision building (Cho & Jonassen, 2002): the characters resituate the choice taken using a simulated dialogue that follows a line of argument until a new decision is made.

If the objective of the role play were to be made explicit from the outset, such as "always use a condom in a risk situation," it would probably lose all interest, in particular for young people. However, if we had to teach skills that had been previously agreed upon with adult subjects, the choice of a scenario based on explicit objectives would be a more recommendable option. Yet, one of the characteristics of games that seek to simulate real situations is that the player does not always know the objective of the game, at least the first time it is played. This gives rise to a certain ambiguity between the objectives of the instructional design (which include a modeling of behavior in risk situations, as well as a set of negotiation skills via the choices to be made and the story that unfolds between the decisions) and those of the player playing the role game for the first time, who does not know very well what it comprises—identify with one of the characters, accompany him or her throughout the journey, and make decisions along the way—but without an explicit objective as to the target that needs to be reached. This ambiguity, or lack of definition as far as the player is concerned, results in a much more situated performance, as the player plays "as if" he or she were one of the characters, making the decisions that they consider to be "normal" whenever required to do so. If the game leads to the player contracting the infection because he or she has engaged in unsafe sexual practices, this simply emphasizes the need for reflection on these practices, reminding the player when and how the behavior occurred but giving the player the opportunity to make mistakes.

The Question of Feedback

Furthermore, as discussed earlier, in the case of AIDS, there is a marked time lag between the occurrence of the risk behavior and the realization of its consequences, which means it is not possible to offer immediate corrective feedback (which would be the most efficient way of doing so).

Several techniques are available for making subjects aware of the anticipated effects of their behavior: one such technique is that used by González (1995), who constructs a graphic simulator of the spread of an infection in a given population (a discotheque) based on sexual relation profiles. The simulator clearly shows how the infection spreads over time and its consequences in the medium and long terms. This technique is very useful for showing the effects on populations or groups and for understanding the epidemic nature of many diseases.

A further didactic technique for individual use, and which was in fact analyzed for use in this project, is that which is used in the application: ***If you love me, show me*** (Family of the Americas, 1995). Although it does not deal directly with AIDS but rather with sexual relations among young people, it uses an animated narrative to show the ups and downs in a date between a boyfriend and girlfriend. Told from the perspective of the girl, it adopts a highly conservative ideology. This undermines somewhat the veracity of the situation described, and the fact that it has few opportunities for interaction means that it is not of direct interest for us. However, it does introduce a form of behavior modeling that might be considered a form of anticipatory feedback: an imaginary character acts as "the conscience" of the teenage girl, pointing out to her when having to make a decision the hidden intentions of the male character and the consequences of a particular action. What is surprising about the technique is that it operates as an unsolicited source of help before the action occurs.

Both techniques are responses to very different approaches but are not especially applicable to the design of our project: in the first case, the simulator is applied to a group in which the profiles of sexual behaviour determine the consequences. It is not possible for subjects to place themselves within a group and experience the evolution for themselves, as the situation does not depend on their own decision-making skills. Furthermore, the behavior profiles are not necessarily recognizable by the individuals as being their own. In the second case, it is virtually impossible for the user to commit mistakes given its preventive nature, with the result that the type of learning is unlikely to be integrated within the subject's action schema. Schank (1999), following a long tradition of "active pedagogy," insists, rightly we believe, in the need to make mistakes and then to rectify these errors so that the actions committed become true learning experiences and modify our prior schema or scripts.

The solution adopted here is for the role play to reduce the period of real time (in fact, the so-called "window period" extends from three to six months), by using a time step that is resolved in the final situation. This means of representing the passing of time allows the player to receive delayed feedback, though it is in fact given during the same session in which the game is played. Given that the success of the game depends on it being as realistic as possible, this time difference does not have any major effect on the game's realism: first, it is a typical technique of audiovisual media and the language of the cinema; and, second, the final situation is not interactive and is used to reveal to each player the results of the game, depending on the character he or she has adopted. This technique allows us, albeit with some limitations, to overcome the problems of the apparent lack of connection between an action and its delayed consequences, as well as to be able to give feedback on the choices made in a very brief period of time (in the real game time). Furthermore, it does away with the need to introduce other solutions such as those mentioned, which would lead to interactions that are not always cohesive with the educational objectives.

However, any application such as the one analyzed here raises many unanswered questions. To what extent are the skills actually learned? Can we really speak of changes in attitude? Are the skills and attitudinal changes transferred to other situations? Clearly, it is not possible to answer

these questions directly, as we are dealing with complex skills. What is required is a longitudinal study, which in this case, as in others before, has been ruled out.

CONCLUSION

Role-playing games are complex multimedia applications, not just because of the technology they use but also because of the way in which this technology responds to an educational framework design that combines a constructivist approach with other concepts, such as the sociocultural modeling of the action and dialogues of argumentation. It would seem that some of the most recurring and justified criticisms that have been made of educational multimedia packages—namely, their technological and multimedia excesses, combined with a lack of pedagogic underpinnings—can be overcome when the educational design is central to the project and the technologies and the production of various multimedia elements are put at the service of the project.

Perhaps the most general conclusion is that each project requires a specific educational design. Just as different subjects are taught in different ways, multimedia applications should be far more specific in their instructional designs, whether their theoretical backgrounds are cognitivist or constructivist.

Equally, it does not appear that the changes in attitude or the transfers of skills are solely attributable to the use of multimedia tools, however complex or well designed they might be. Rather, they would seem to be the result of more than *one* educational action. From the perspective of health education and, in particular from that of AIDS prevention in the young, it would appear that multimedia role-playing games need to be complemented with more formal instruction techniques backed with a range of additional activities.

ACKNOWLEDGMENT

The project was funded by the EU programme "'Europe against AIDS," and was a joint undertaking between the cooperative group, CLAPS, of Pordenone, Italy, and the University of Barcelona (ICE), Spain. The author wishes to express his gratitude to Carlo Mayer, co-director on behalf of CLAPS, and to the Spanish team of Begoña Gros, Cristina Martínez, and María José Rubio.

REFERENCES

Andreu, O. A. (1991). Sida y Antropologia social, en *Jano*, Marzo, 1(942), 51.

Bandura, A. (1987). *Pensamiento y acción*. Barcelona: Martínez Roca.

Bayés, R. (1987). Factores de aprendizaje en salud y enfermedad. *Revista Española de Terapia del Comportamiento, 5*(2), 119–135.

Bayés, R. (1994). *Sida i Psicologia*. Barcelona: Martínez Roca.

Brooks-Gunn, J., Boyer, C. B., & Hein K. (1988). Preventing HIV infection and AIDS in children and adolescent. *American Psychologist*, November, *1*(11), 958–964.

Cho, K. L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology: Research & Development, 50*(3), 5–22.

Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 170–198). New York: Simon & Schuster Macmillan.

Duffy, T. M., Lowyck, J., & Jonassen, D. H. (Eds.). (1992). *Designing environments for constructivist learning.* Heidelberg: Springer.

Fundació LaCaixa. (1995). *Sida. Saber ayuda.* Barcelona: La Caixa.

Gee, J. P. (2003). *What videogames have to teach us about learning and literacy*? New York: Palgrave-MacMillan.

Gonzalez, J. J. (1995). Computer assisted learning to prevent HIV spread: Visions, delays and opportunities. *Machine-Mediated Learning, 5*(1), 3–11.

Green, L. W., Kreute, M. W., Deedds, S. G., & Partridge, K. B. (1980). *Health education planning: A diagnostic approach.* Palo Alto, CA: Mayfield.

Hill, B. (2000). *The magic of reading.* Redmon: Microsoft.

Jonassen, D., Peck, K., & Wilson, B. (1999). *Learning with technology. A constructivist perspective.* Upper Saddle River, NJ: Prentice-Hall.

Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

Lave, J. (1988). *La cognición en la práctica.* Barcelona: Paidós.

Lave, J. (1990). The culture of acquisition and the practice of understanding. In D. Kirshner & J. A. Whitson (Eds.), *Situated cognition* (pp. 17–36). Mahwah, NJ: Lawrence Erlbaum Associates.

Plowman, L. (1994). The "Primitive Mode of Representation" and the evolution of interactive multimedia. *Journal of Educational Multimedia and Hypermedia, 3*(3/4), 275–293.

Rodríguez Illera, J. L., Gros, B., Martínez, C., & Rubio, M. J. (1999). Un software multimedia para la prevención del SIDA en adolescentes. *Multimedia educativo 99.* Barcelona: Universitat de Barcelona.

Schank, R. C. (1998). *Inside multi-media case based instruction.* Hillsdale, NJ: Erlbaum.

Schank, R. C. (1999). *Dynamic memory revisited.* Cambridge, MA: Cambridge University Press.

Tonks, D. (1996). *Teaching aids.* New York: Routledge.

Wenger, E. (1998). *Communities of practice.* Cambridge, MA: Cambridge University Press.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, pp. 271-288, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.14
Location-Based Multimedia Services for Tourists

Panagiotis Kalliaras
National Technical University of Athens, Greece

Athanasios-Dimitrios Sotiriou
National Technical University of Athens, Greece

P. Papageorgiou
National Technical University of Athens, Greece

S. Zoi
National Technical University of Athens, Greece

INTRODUCTION

The evolution of mobile technologies and their convergence with the Internet enable the development of interactive services targeting users with heterogeneous devices and network infrastructures (Wang et al., 2004). Specifically, as far as cultural heritage and tourism are concerned, several systems offering location-based multimedia services through mobile computing and multimodal interaction have already appeared in the European research community (LOVEUS, n.d.; Karigiannis, Vlahakis, & Daehne, n.d.).

Although such services introduce new business opportunities for both the mobile market and the tourism sector, they are not still widely deployed, as several research issues have not been resolved yet, and also available technologies and tools are not mature enough to meet end user requirements. Furthermore, user heterogeneity stemming both from different device and network technologies is another open issue, as different versions of the multimedia content are often required.

This article presents the AVATON system. AVATON aims at providing citizens with ubiquitous user-friendly services, offering personalized, location-aware (GSM Association, 2003), tourism-oriented multimedia information related to the area of the Aegean Volcanic Arc. Towards this end, a uniform architecture is adopted in

order to dynamically release the geographic and multimedia content to the end users through enhanced application and network interfaces, targeting different device technologies (mobile phones, PDAs, PCs, and TV sets). Advanced positioning techniques are applied for those mobile user terminals that support them.

SERVICES

AVATON is an ambient information system that offers an interactive tour to the user (visitor) in the area of the Aegean Volcanic Arch (see http://www.aegean.gr/petrified_forest/). The system can serve both as a remote and as an onsite assistant for the visitor, by providing multimedia-rich content through various devices and channels:

- Over the Internet, via Web browsers with the use of new technologies such as rich-clients and multi-tier architecture in order to dynamically provide the content;
- With portable devices (palmtops, PDAs) and 2.5G or 3G mobile phones, which are capable of processing and presenting real-time information relevant to the user's physical position or areas of interest; and
- Via television channels—AVATON allows users to directly correlate geographic with informative space and conceivably pass from one space to the other, in the context of Worldboard (Spohrer, 1999).

With the use of portable devices equipped with positioning capabilities, the system provides:

- Dynamic search for geographical content, information related to users' location, or objects of interest that are in their proximity;
- Tours in areas of interest with the aid of interactive maps and 3-D representations of the embossed geography;
- Search for hypermedia information relative to various geographic objects of the map;
- User registration and management of personal notes during the tour that can be recalled and reused during later times; and
- Interrelation of personal information with natural areas or objects for personal use or even as a collective memory relative to visited areas or objects.

THE AVATON ARCHITECTURE

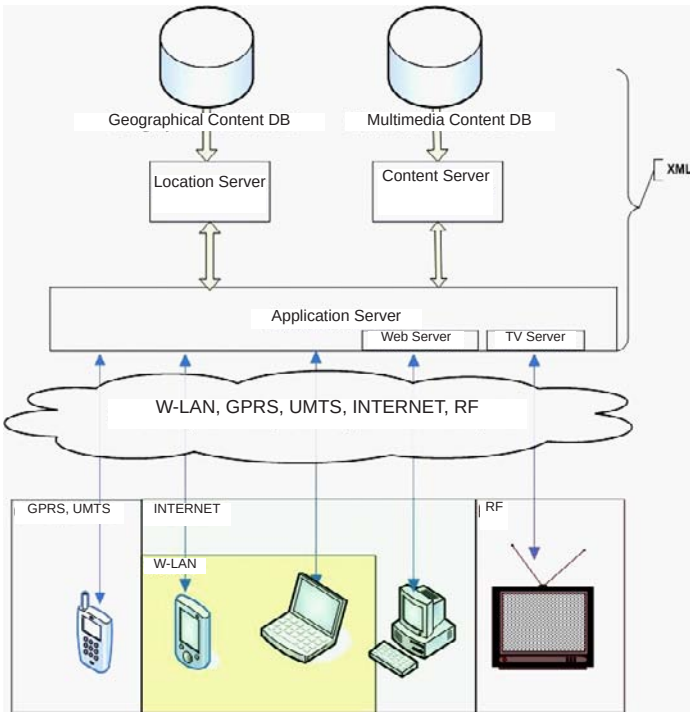
Overview

The AVATON system is based on a client-server architecture composed of three main server components: the application server, the content server, and the location server. The application server combines information and content from the content and location servers, and replies to client requests through different network technologies. The content is retrieved from two kinds of databases, the geographical and multimedia content DBs. The above architecture is shown in Figure 1.

In more detail:

- **Multimedia Content Database:** This database contains the multimedia content such as images, video, audio, and animation.
- **Geographical Content Database:** A repository of geographical content such as aerial photos, high-resolution maps, and relevant metadata.
- **Content Server:** The content server supplies the application server with multimedia content. It retrieves needed data from the multimedia content database according to user criteria and device capabilities, and responds to the application server.
- **Location Server:** Serves requests for geographical content from the application server by querying the geographical content data-

Figure 1. The AVATON architecture



base. The content retrieved is transformed into the appropriate format according to user device display capabilities and network bandwidth available.

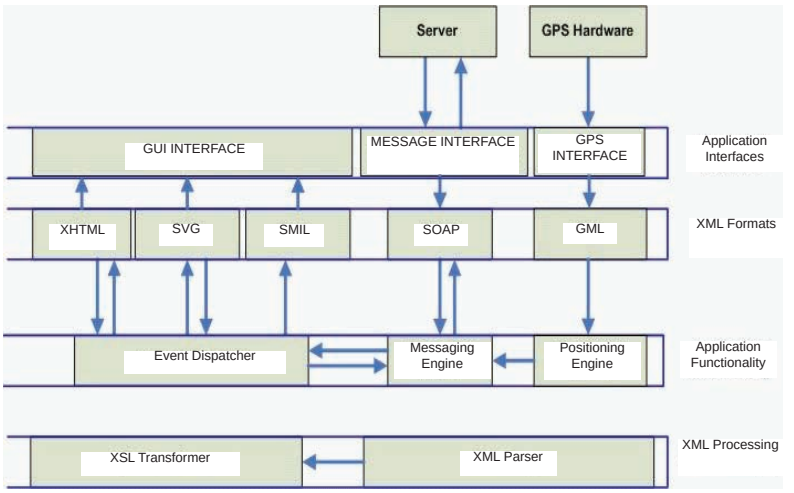
- **Application Server:** The application server receives requests from different devices through GPRS, UMTS (third-generation mobile phone), W-LAN, Internet (PDA, laptop, PC), and RF (television). The server identifies each device and transmits data in an appropriate format. More precisely, the application server incorporates a Web server and a TV server in order to communicate with PCs and televisions respectively.

Client

This section focuses on the mobile-phone and PDA applications. The scope of the AVATON system includes Java-enabled phones with color displays and PDAs with WLAN or GPRS/UMTS connectivity. While all the available data for the application can be downloaded and streamed over the network, data caching is exploited for better performance and more modest network usage.

When the users complete their registration in the system, they have in their disposal an interactive map that initially portrays the entire region as well as areas or individual points of interest. For acquiring user position, the system is using GPS. The client also supports multi-lingual implementation, as far as operational content is concerned, for example menus, messages, and help. These files are maintained as XML documents. XML is extensively used in order to ease the load of parsing different data syntaxes. A single process, the XML Parser is used for decoding all kinds of data and an XSL Transformer for transcoding them in new formats. The different XML formats are XHTML, SVG, SMIL, SOAP, and GML, as shown in Figure 2.

Figure 2. The XML-based technologies in the client side



Geographical Info Presentation

In order to render the geographical data, the client receives raster images for the drawing of the background map, combined with metadata concerning areas of interest and links to additional textual or multimedia information. The raster data are aerial high-resolution photographs of the region on two or three scales. Because of the high resolution of the original images, the client is receiving small portions, in the form of tiles from the *raster data processing engine* in the server side, which are used to regenerate the photorealistic *image layer* in a resolution that is suitable for the device used. The attributes of the geographical data are generated in vector ShapeFile (ESRI ShapeFile) format, which is quite satisfactory for the server side but not for lightweight client devices. So, a *SHP TO SVG converter* at the server side is regenerating the metadata in SVG format that can be viewed properly from a handheld device. As soon as the metadata is downloaded to the client device, a final filtering (XSLT transformation) is done and the additional layer is opposed to the image layer in the *SVG viewer.* On the *SVG data layer*, the user can interact with points of interest and receive additional information in the form of text or multimedia objects. The above are shown in Figure 3.

Multimedia Info Presentation

The presentation of multimedia information mainly depends on user position. The system is designed to provide audio and video clips, 3-D representations, and also textual information concerning each place of interest. Not all devices, though, receive the same content, since they differ in display, processor, or network speed. For that purpose, for each registered user, the system decides what kind of content is more suitable for them to receive and the multimedia content server generates the appropriate script. Depending on the available memory of the client's device, media objects stay resident in the cache memory so that frequently requested content is accessed without delays that occur due to network latency. In Figure 4 the components that are involved in the multimedia presentation are shown. The *TourScript Data* contains the script which describes the multimedia presentation. It is transcoded inside the *SMIL generator* to a SMIL message that follows the XML syntax, so that it can be incorporated seamlessly to the messages

Figure 3. Client-side map rendering

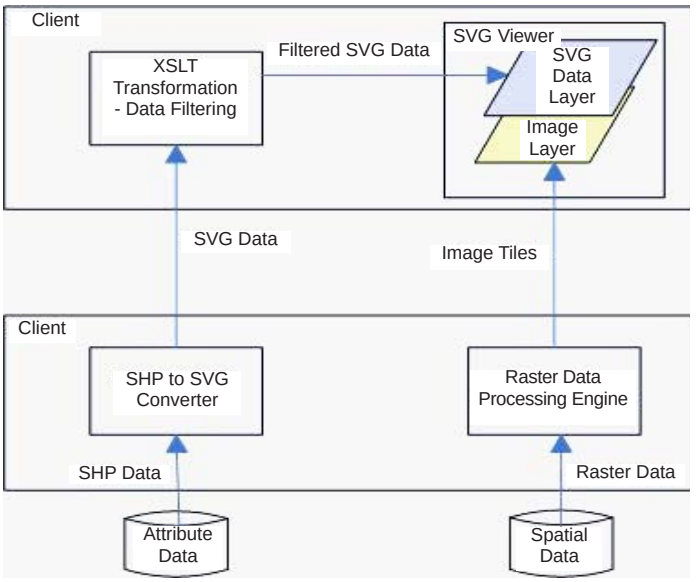
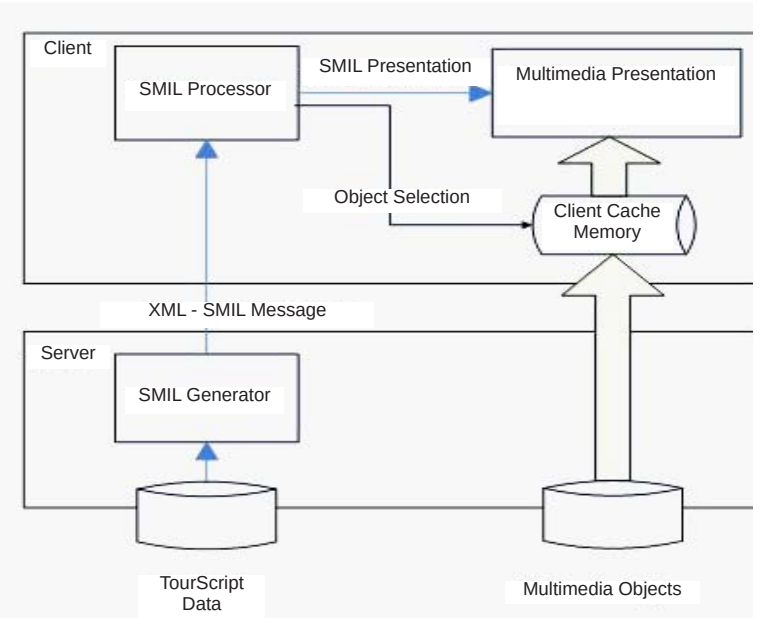


Figure 4. Client-side map rendering



that are exchanged in the AVATON system. At the client side, the SMIL message is received by the *SMIL processor* which coordinates the process of fetching the *multimedia objects* from the *client cache memory* to the suitable renderer, so that the *multimedia presentation* can be completed.

Location Server

The location server is the component that handles the geographical content of the AVATON system. It provides a storage system for all geographical data and allows querying of its contents through location criteria, such as global position and areas of interest. Content management is based on a PostgreSQL (http://www.postgres.org) relational database. A JDBCInterface uses the JDBC APIs in order to provide support for data operations. A GISExtension is also present, based on PostGIS (http://www.postgis.org), in order to enable the PostgreSQL server to allow spatial queries. This feature is utilized through a GISJDBCInterface, a PostGIS layer on top of PostgreSQL.

Cartographic Data

Concerning the photorealistic information, the user can choose from several distinct zoom levels. The mobiles phones and PDAs in the market that support GPRS or WLAN have displays of different resolutions that, in most cases, are multiples of 16 pixels. Hence, the location server can generate tiles with a multiple of 16p x 16p, which can be presented in the user's mobile device. The server always holds multiple resolutions for every level of cartographic (photographic) information. The levels of cartographic information define the degree of focus.

Apart from the photographic layer, additional layers of vector information also exist, and their size is approximately 5% or 10% of the corresponding photographic. Therefore, in practice, every device will initially request from the server the cartographical information with the maximum resolution this device can support. Hence, the server decides the available resolution that best corresponds to the requested resolution from the device. The size of every tile is approximately 1K. The devices with greater resolution per tile receive more files, with greater magnitude, for every degree of focus due to the higher resolution.

Multimedia Content Server

The multimedia content server component comprises the major unit that controls the mixing and presentation of different multimedia objects. Its purpose is to upload all the objects necessary and present them in a well-defined controlled order that in general depends on the user position, interactions, and tracking information available. The multi-lingual audiovisual information scheduled for presentation is coordinated so that several objects may be presented simultaneously. The multimedia content server component is also responsible for choosing "relevant" objects for the user to select among in the case the user requires more information on a topic. The multimedia content server interfaces with the multimedia content database, a relational database storing the multimedia content. The database is organized thematically and allows the creation of hierarchical structures. It also contains a complete list of multimedia material, covering all content of the physical site, such as 3D reconstructed plants, audio narration, virtual 3D models, avatar animations, and 2D images.

Media Objects

As mentioned already, the multimedia content server is responsible for mixing the basic units of multimedia information. These elements are hierarchically ordered. At the finest level of granularity, there are atomic objects called MediaObjects with specializations such as AudioMediaObject, ImageMediaObject, 3DMediaObject, and CompositeMediaObject. These objects contain the

actual data to be rendered along with additional profile metadata characterizing them. At a higher level of complexity, a TourScript represents an ordered sequence of MediaObjects, all of which are to be presented if the script is chosen.

According to user requirements, the user will be able to navigate through the site in a geographically based tree. This is made possible through the use of points of interest (PoI) and areas of interest (AoI). A PoI can only contain TourScripts and can be viewed as the end node of the site tree. In contrast, an AoI may contain either another PoI, an AoI, or TourScripts. This allows the system to map the actual site into a hierarchy model containing PoI at the top and MediaObject components at the leaf level.

The multimedia content server is also responsible for managing this site-tree for the entire site. Moreover it is responsible for traversing it. The use of the site tree is quite interesting: when a media object, for instance *audio* object, is presented to the user, it belongs to a node in the site hierarchy. Figure 5 shows the structure of the site in a tree view as described previously. The multimedia content server is responsible for coordinating the rendering components in order to provide a synchronized presentation to the user, according to user preferences, position, and commands.

Deployment and Usage

Based on the proposed architecture, the AVATON services are being deployed to physical sites within the Aegean Volcano Arc (such as Santorini and Lesvos islands) and evaluated by real end users under different scenarios. The main air interfaces that will be used by the system (along with standard wired access through common LANs or the Internet) are:

- **WLAN:** A standard 802.11b wireless access network provides connectivity for users equipped with portable devices (such as PDA with built-in WLAN cards or laptops).
- **GPRS/UMTS:** For mobile users and smart phones, access will be provided through the GSM network, using GPRS (Bettsteter, Vögel, & Eberspächer, 1999). This restricts the system from providing video or 3-D animations to such users, and the services offered are focused on text, images (including map information), and short audio. As GPRS is already packet oriented, our implementation can be easily transferred to UMTS, if available.

Figure 5. Hierarchy formulation of media objects at the multimedia content server

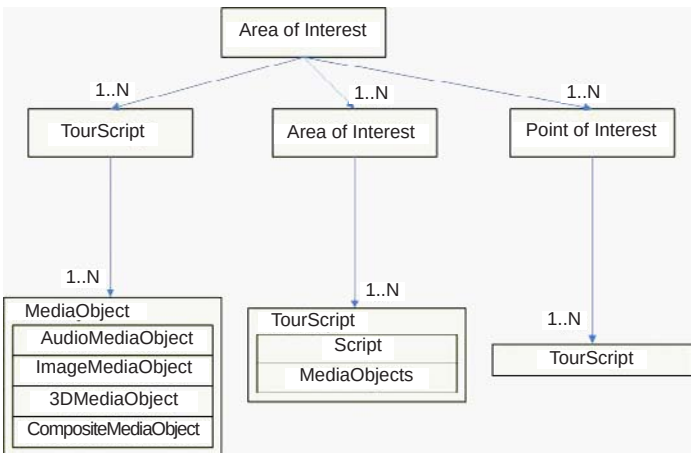


Figure 6. Implementation plan for the Lesvos island site

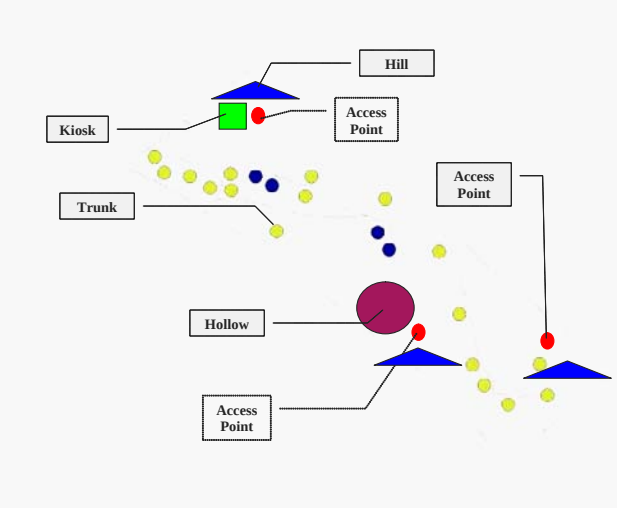
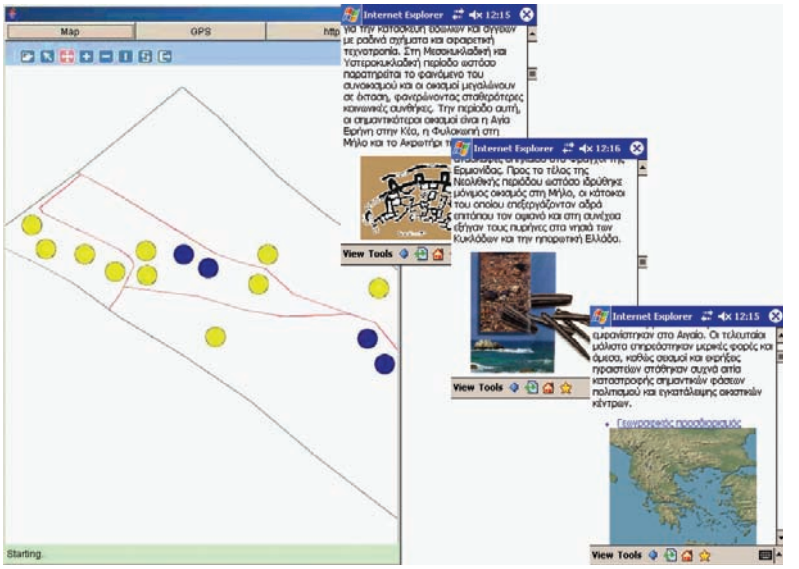


Figure 7. SVG map and interface



In Figure 6, the actual implementation plan is given for the location of the Sigri Natural History Museum on Lesvos island. The area consists of an open geological site, the Petrified Forest, where the ash from a volcanic eruption some 15 to 20 million years ago covered the stand of sequoia trees, causing their petrification. Wireless access is provided by the use of a 3 Netgear 54Mbps access point equipped with additional Netgear antennas in order to overcome the physical limitations of the area (hills, trunks, and hollows).

Visitors to the site are equipped with PDAs or smart phones (provided at the entrance kiosk) and stroll around the area. A typical scenario consists of the following: The users enter the archaeological site and activate their devices. They then perform

a login and provide personal details to the server, such as username, language selection, and device settings. The client then requests from the server and loads the map of the area in SVG (http://www.w3.org/TR/SVG) format, as seen in Figure 7. The circles on the map present distinct points of interest (yellow indicating trees and blue indicating leafs). The application monitors the users' location and updates the SVG map in real time, informing the users of their position. The SVG map is interactive, and when the users enter the vicinity of a point of interest, the application automatically fetches and displays (via their browser or media player) the corresponding multimedia content (in the form of HTML pages, audio, or video) at the requested language. The users are also able to navigate manually through the available content and receive additional information on topics of their interest. During the tour, the users' path is being tracked and displayed (the red line on Figure 7) in order to guide them through the site. They are also able to keep notes or mark favorite content (such as images), which can be later sent to them when they complete the tour.

FURTHER WORK

The system is currently being deployed and tested in two archeological sites. Users are expected to provide useful feedback on system capabilities and assist in further enhancements of its functionality. Also, as 3G infrastructure is being expanded, incorporation of the UMTS network in the system's access mechanisms will provide further capabilities for smart phone devices and also use of the system in areas where wireless access cannot be provided.

Towards commercial exploitation, billing and accounting functionalities will be incorporated into the proposed architecture. Finally, possible extensions of the system are considered in order to include other cultural or archeological areas.

REFERENCES

Bettsteter, C., Vögel, H.-J., & Eberspächer, J. (1999). GSM Phase 2+ General Packet Radio Service GPRS: Architecture, protocols, and air interface. *IEEE Communication Surveys,* (3rd Quarter).

ESRI ShapeFile Technical Description. (1998, July). *An ESRI white paper.*

GSM Association. (2003, January). Location based services. *SE.23, 3.10.*

Karigiannis, J. N., Vlahakis, V., & Daehne, P. (n.d.). ARCHEOGUIDE: Challenges and solutions of a personalized augmented reality guide for archeological sites. *Computer Graphics in Art, History and Archeology, Special Issue of the IEEE Computer Graphics and Application Magazine.*

LOVEUS. (n.d.). Retrieved from http://loveus.intranet.gr/documentation.htm

Spohrer, J. C. (1999). Information in places. *IBM System Journal, 38*(4).

Wang, Y., Cuthbert, L., Mullany, F. J., Stathopoulos, P., Tountopoulos, V., Sotiriou, D. A., Mitrou, N., & Senis, M. (2004). Exploring agent-based wireless business models and decision support applications in an airport environment. *Journal of Telecommunications and Information Technology,* (3).

KEY TERMS

Cartographic Data: Spatial data and associated attributes used by a geographic information system (GIS).

Content Provider: A service that provides multimedia content.

Location-Based Services: A way to send custom advertising and other information to cell phone subscribers based on their curent location.

Multimedia: Media that usesmultiple forms of information content and information processing (e.g., text, audio, graphics, animation, video, interactivity) to inform or entertain the (user) audience.

Network: A network of telecommunications links arranged so that data may be passed from one part of the network to another over multiple links.

Tourism: The act of travel for predominantly recreational or leisure purposes, and the provision of services in support of this act.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 387-392, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.15
Software Engineering for Mobile Multimedia
A Roadmap

Ghita Kouadri Mostéfaoui
University of Fribourg, Switzerland

ABSTRACT

The abstract should be changed to this new abstract: Research on mobile multimedia mainly focuses on improving wireless protocols in order to improve the quality of services. In this chapter, we argue that another perspective should be investigated in more depth in order to boost the mobile multimedia industry. This perspective is software engineering which we believe it will speed up the development of mobile multimedia applications by enforcing reusability, maintenance, and testability. Without any pretense of being comprehensive in its coverage, this chapter identifies important software engineering implications of this technological wave and puts forth the main challenges and opportunities for the software engineering community.

INTRODUCTION

A recent study by Nokia (2005) states that about 2.2 billion of us are already telephone subscribers, with mobile subscribers now accounting for 1.2 billion of these. Additionally, it has taken little more than a decade for mobile subscriptions to outstrip fixed lines, but this still leaves more than half the world's population without any kind of telecommunication service. The study states that this market represents a big opportunity for the mobile multimedia industry.

Research on mobile multimedia mainly focuses on improving wireless protocols in order to improve the quality of service. In this chapter, we argue that another perspective should be investigated in more depth in order to boost the mobile multimedia industry. This perspective is software engineering which we believe it will speed up the development of mobile multimedia applications by enforcing reusability, maintenance, and

testability of mobile multimedia applications. Without any pretense of being comprehensive in its coverage, this chapter identifies important software engineering implications of this technological wave and puts forth the main challenges and opportunities for the software engineering community.

ORGANIZATION OF THIS CHAPTER

The next Section presents the state of the art of research in mobile multimedia. The section "What Software Engineering Offers to Mobile Multimedia?" argues on the need for software engineering for mobile multimedia. The section "Contributions to 'Mobile' Multimedia Software Engineering" surveys initiatives in using software engineering techniques for the development of mobile multimedia applications. The section "Challenges of Mobile Multimedia Software Engineeering " highlights the main challenges of mobile multimedia software engineering. Some of our recommendations for successfully bridging the gap between software engineering and mobile multimedia development are presented. The last section concludes this chapter.

STATE OF THE ART OF CURRENT RESEARCH IN MOBILE MULTIMEDIA

I remember when our teacher of "technical terms" in my Engineering School introduced the term "Multimedia" in the middle of the 1990s. He was explaining the benefits of Multimedia applications and how future PCs will integrate such capabilities as a core part of their design. At this time, it took me a bit before I could understand what he meant by integrating image and sound for improving user's interactivity with computer systems. In fact, it was only clear for me when I bought my first "Multimedia PC."

Multimedia is recognized as one of the most important keywords in the computer field in the 1990s. Initially, communication engineers have been very active in developing multimedia systems since image and sound constitute the langua franca for communicating ideas and information using computer systems through networks. The broad adoption of the World Wide Web encouraged the development of such applications which spreads to other domains such as remote teaching, e-healthcare, and advertisement. People other than communication engineers have also been interested in multimedia like medical doctors, artists, and people in computer fields such as databases and operating systems (Hirakawa, 1999).

Mobile multimedia followed as a logical step towards the convergence of mobile technologies and multimedia applications. It has been encouraged by the great progress in wireless technologies, compression techniques, and the wide adoption of mobile devices. Mobile multimedia services promote the realization of the ubiquitous computing paradigm for providing anytime, anywhere multimedia content to mobile users. The need for such content is justified by the huge demand for a quick and concise form of communication–compared to text—formatted as an image or an audio/video file. A recent study driven by MORI, a UK-based market researcher (LeClaire, 2005), states that the demand for mobile multimedia services is on the rise, and that the adoption of mobile multimedia services is set to take off in the coming years and will drive new form factors. The same study states that 90 million mobile phones users in Great Britain, Germany, Singapore, and the United States, are likely to use interactive mobile multimedia services in the next two years.

We are looking at the cell phone as the next big thing that enables mobile computing, mainly because phones are getting smarter" Burton Group senior analyst Mike Disabato told the E-Commerce Times. "We'll see bigger form factors coming in some way, shape or form over the next

few years. Those form factors will be driven by the applications that people want to run.

In order to satisfy such a huge demand, research has been very active in improving current multimedia applications and in developing new ones driven by consumers' needs, such as mobile IM (Instant Messaging), group communication, and gaming, along with speed and ease of use. When reviewing efforts in research on mobile multimedia, one can observe that most of the contributions fall into the improvement of wireless protocols and development of new mobile applications.

Mobile Networks

Research on wireless protocols aims at boosting mobile networks and Internet to converge towards a series of steps:

- **WAP:** In order to allow the transmission of multimedia content to mobile devices with a good quality/speed ratio, a set of protocols have been developed and some of them have been already adopted. The wireless application protocols (WAP), aim is the easy delivery of Internet content to mobile devices over GSM (global system for mobile communications), is published by the WAP Forum, founded in 1997 by Ericsson, Motorola, Nokia, and Unwired Planet. The WAP protocol is the leading standard for information services on wireless terminals like digital mobile phones and is based on Internet standards (HTML, XML, and TCP/IP). In order to be accessible to WAP-enabled browsers, Web pages should be developed using WML (Wireless Markup Language), a mark-up language based on XML and inherited from HTML.
- **GPRS:** The General Packet Radio Service is a new non-voice value added service that allows information to be sent and received across a mobile telephone network (GSM World, 2005). GPRS has been designed to facilitate several new applications that require high speed such as collaborative working, Web browsing, and remote LAN access. GPRS boosts data rates over GSM to 30-40 Kbits/s in the packet mode.
- **EDGE:** The Enhanced Data rates for GSM Evolution technology is an add-on to GPRS and therefore cannot work alone. The EDGE technology is a method to increase the data rates on the radio link for GSM. It introduces a new modulation technique and new channel coding that can be used to transmit both packet-switched and circuit-switched voice and data services (Ericsson, 2005). It enjoys a data rate of up 120-150 Kbits/s in packet mode.
- **UMTS:** Universal Mobile Telecommunications Service is a third-generation (3G) broadband, packet-based transmission of text, digitized voice, video, and multimedia at data rates up to 2 megabits per second (Mbps) that offers a consistent set of services to mobile computer and phone users no matter where they are located in the world (UMTS, 2005).

Research on wireless protocols is still an active field supported by both academia and leading industry markets.

Mobile Multimedia Applications

With the advantages brought by third-generation (3G) networks like the large bandwidth, there are many chances that PDAs and mobile phones will become more popular than PCs since they will offer the same services with mobility as an added-value. Jain (2001) points out that important area where we can contribute important ideas is in improving the user's experience by identifying the relevant applications and technology for mobile multimedia.

Currently, the development of multimedia applications for mobile users is becoming an active field of research. This trend is encouraged by the high demand of such applications by mobile users from different fields of applications ranging from gaming, rich-information delivery, and emergencies management.

WHAT SOFTWARE ENGINEERING OFFERS TO MOBILE MULTIMEDIA?

Many courses on software engineering multimedia are taught all over the world. Depicting the content of these courses shows a great focus on the use of multimedia APIs for human visual system, signal digitization, signal compression, and decompression. Our contribution, rather, falls into software engineering in its broader sense including software models and methodologies.

Multimedia for Software Engineering vs. Software Engineering for Multimedia

Multimedia software engineering can be seen in two different, yet complementary roles:

1. The use of multimedia tools to leverage software engineering
2. The use of software engineering methodologies to improve multimedia applications development

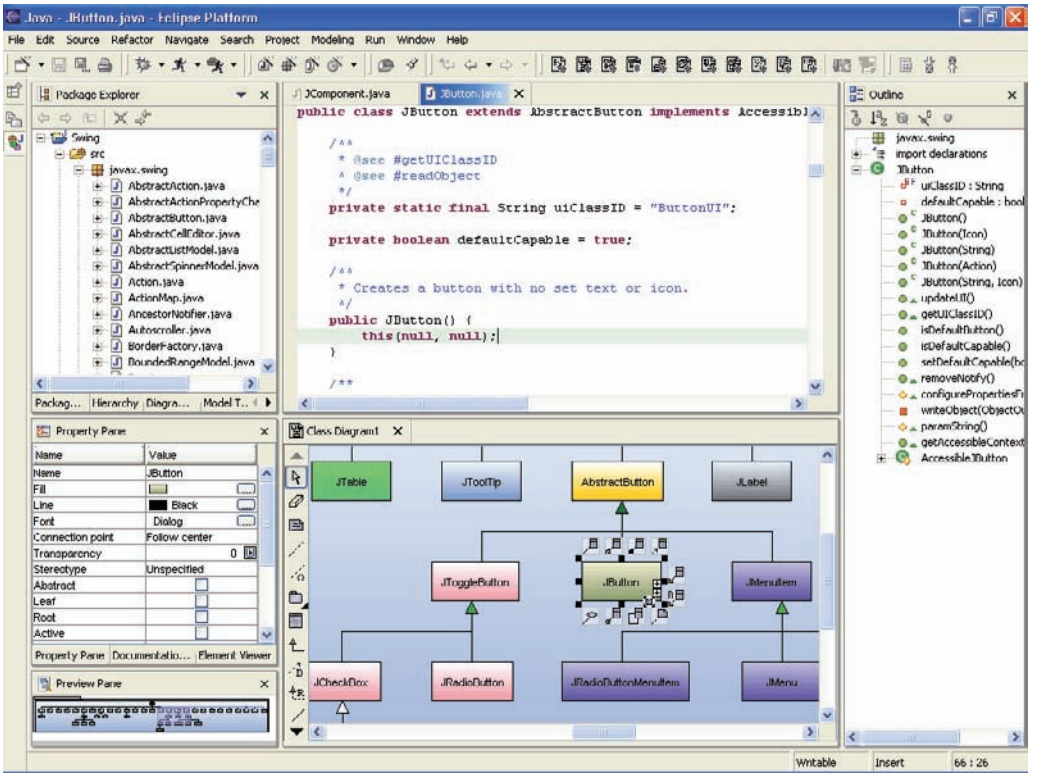
Examples of the first research trail are visual languages and software visualization. Software Visualization aims at using graphics, pretty-printing, and animation techniques to show program code, data, and dependencies between classes and packages. Eclipse (Figure 1), TogetherSoft, and Netbeans are example tools that use multimedia to enhance code exploration and comprehension.

The second research trail is a more recent trend and aims at improving multimedia software development by relying on the software engineering discipline. An interesting paper by Masahito Hirakawa (1999) states that software engineers do not seem interested in multimedia. His guess is that "they assume multimedia applications are rather smaller than the applications that software engineers have traditionally treated, and consider multimedia applications to be a research target worth little." He argues that the difference between multimedia and traditional applications is not just in size but also the domain of application. While there is no disagreement on this guess, it would be more appropriate to expand. We claim that there is a lack of a systematic study that highlights the benefits of software engineering for multimedia. Additionally, such study should lay down the main software approaches that may be extended and/or customized to fit within the requirements of "mobile" multimedia development.

Due to the huge demand of software applications by the industry, the U.S. President's Information Technology Advisory Committee (PITAC) report puts "Software" as the first four priority areas for long-term R&D. Indeed, driven by market pressure and budget constraints, software development is characterized by the preponderance of ad-hoc development approaches. Developers don't take time to investigate methodologies that may accelerate software development because learning these tools and methodologies itself requires time. As a result, software applications are very difficult to maintain and reuse, and most of the time related applications-domains are developed from scratch across groups, and in the worst case in the same group.

The demand for complex, distributed multimedia software is rising; moreover, multimedia software development suffers from similar pitfalls discussed earlier. In the next section, we explore the benefits of using software engineering tools and methodologies for mobile multimedia development.

Figure 1. A typical case tool



Software Engineering for Leveraging Mobile Multimedia Development

Even if mobile multimedia applications are diverse in content and form, their development requires handling common libraries for image and voice digitization, compression/decompression, identification of user's location, etc. Standards APIs and code for performing such operations needs to be frequently duplicated across many systems. A systematic reuse of such APIs and code highly reduces development time and coding errors. In addition to the need of reuse techniques, mobile multimedia applications are becoming more and more complex and require formal specification of their requirements. In bridging the gap between software engineering and mobile multimedia, the latter domain will benefit from a set of advantages summarized in the following:

- **Rapid development of mobile multimedia applications:** This issue is of primordial importance for the software multimedia industry. It is supported by reusability techniques in order to save time and cost of development.
- **Separation of concerns:** A mobile multimedia application is a set of functional and non-functional aspects. Examples are security, availability, acceleration, and rendering. In order to enforce the rapid development of applications, these aspects need to be developed and maintained separately.
- **Maintenance:** This aspect is generally seen as an error correction process. In fact, it is broader than that and includes software enhancement, adaptation, and code understanding. That's why, costs related to software maintenance is considerable and

mounting. For example, in USA, annual software maintenance has been estimated to be more than $70 billion. At company-level, for example, Nokia Inc. used about $90 million for preventive Y2K-bug corrections (Koskinen, 2003).

In order to enforce the requirements previously discussed, many techniques are available. The most popular ones are detailed in the next Section including their concrete application for mobile multimedia development.

CONTRIBUTIONS TO "MOBILE" MULTIMEDIA SOFTWARE ENGINEERING

This section explores contributions that rely on software design methodologies to develop mobile multimedia applications. These contributions have been classified following three popular techniques for improving software quality including the ones outlined above. These techniques are: middleware, software frameworks, and design patterns.

Middleware

An accustomed to conferences in computer science has with no doubt attended a debate on the use of the word "middleware." Indeed, it's very common for developers to use this word to describe any software system between two distinct software layers, where in practice; their system does not necessarily obey to middleware requirements.

According to (Schmidt & Buschmann, 2003) middleware is software that can significantly increase reuse by providing readily usable, standard solutions to common programming tasks, such as persistent storage, (de)marshalling, message buffering and queuing, request de-multiplexing, and concurrency control. The use of middleware helps developers to avoid the increasing complexity of the applications and lets them concentrate on the application-specific tasks. In other terms, middleware is a software layer that hides the complexity of OS specific libraries by providing easy tools to handle low-level functionalities.

CORBA (common object request broker architecture), J2EE, and .Net are examples middleware standards that emerge from industry and market leaders. However, they are not suitable for mobile computing and have no support for multimedia.

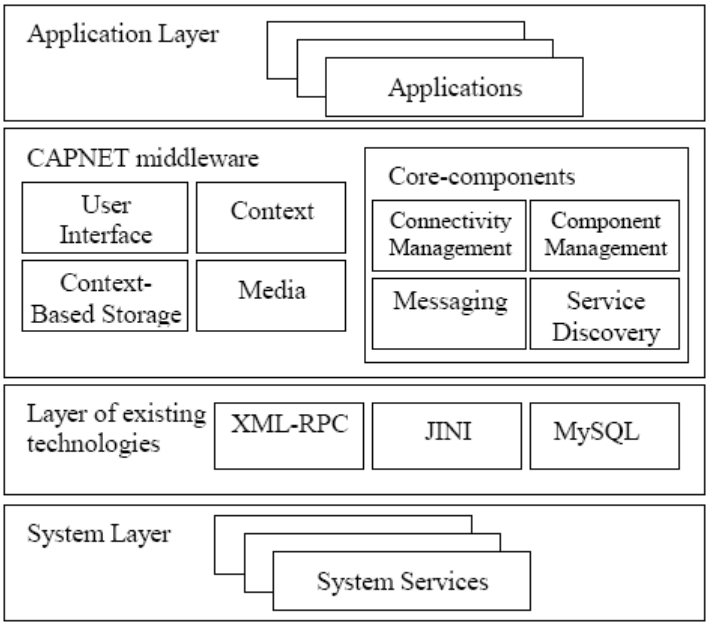
Davidyuk, Riekki, Ville-Mikko, and Sun (2004) describe CAPNET, a context-aware middleware which facilitates development of multimedia applications by handling such functions as capture and rendering, storing, retrieving and adapting of media content to various mobile devices (see Figure 2). It offers functionality for service discovery, asynchronous messaging, publish/subscribe event management, storing and management of context information, building the user interface, and handling the local and network resources.

Mohapatra et al. (2003) propose an integrated power management approach that unifies low level architectural optimizations (CPU, memory, register), OS power-saving mechanisms (dynamic voltage scaling) and adaptive middleware techniques (admission control, optimal transcoding, network traffic regulation) for optimizing user experience for streaming video applications on handheld devices. They used a higher level middleware approach to intercept and doctor the video stream to compliment the architectural optimizations.

Betting on code portability, Tatsuo Nakajima (2002) describes a java-based middleware for networked audio and visual home appliances executed on commodity software. The high-level abstraction provided by the middleware approach makes it easy to implement a variety of applications that require composing a variety of functionalities.

Middleware for multimedia networking is currently a very active area of research and standardization.

Figure 2. The architecture of CAPNET middleware (Davidyuk et al., 2004)



Software Frameworks

Suffering from the same confusion in defining the word middleware, the word "framework" is used to mean different things. However, in this chapter, we refer to frameworks to software layers with specific characteristics we detail in the following. Software frameworks are used to support design reuse in software architectures. A framework is the skeleton of an application that can be customized by an application developer. This skeleton is generally represented by a set of abstract classes. The abstract classes define the core functionality of the framework, which also contains a set of concrete classes that provide a prototype application introduced for completeness. The main characteristics of frameworks are their provision of high level abstraction; in contrast to an application that provides a concrete solution to a concrete problem, a framework is intended to provide a generic solution for a set of related problems. Plus, a framework captures the programming expertise: necessary to solve a particular class of problems. Programmers purchase or reuse frameworks to obtain such problem-solving expertise without having to develop it independently.

Such advantages are exploited in Scherp and Boll (2004) where a generic java-based software framework is developed to support personalized (mobile) multimedia applications for travel and tourism. This contribution provides an efficient, simpler, and cheaper development platform of personalized (mobile) multimedia applications.

The Sesame environment (Coffland & Pimentel, 2003) is another software framework built for the purpose of modeling and simulating heterogeneous embedded multimedia systems.

Even if software frameworks are considered as an independent software technique, they are very often used to leverage middleware development and to realize the layered approach.

Design Patterns

Design patterns are proven design solutions to recurring problems in software engineering. Patterns are the result of developers' experience in

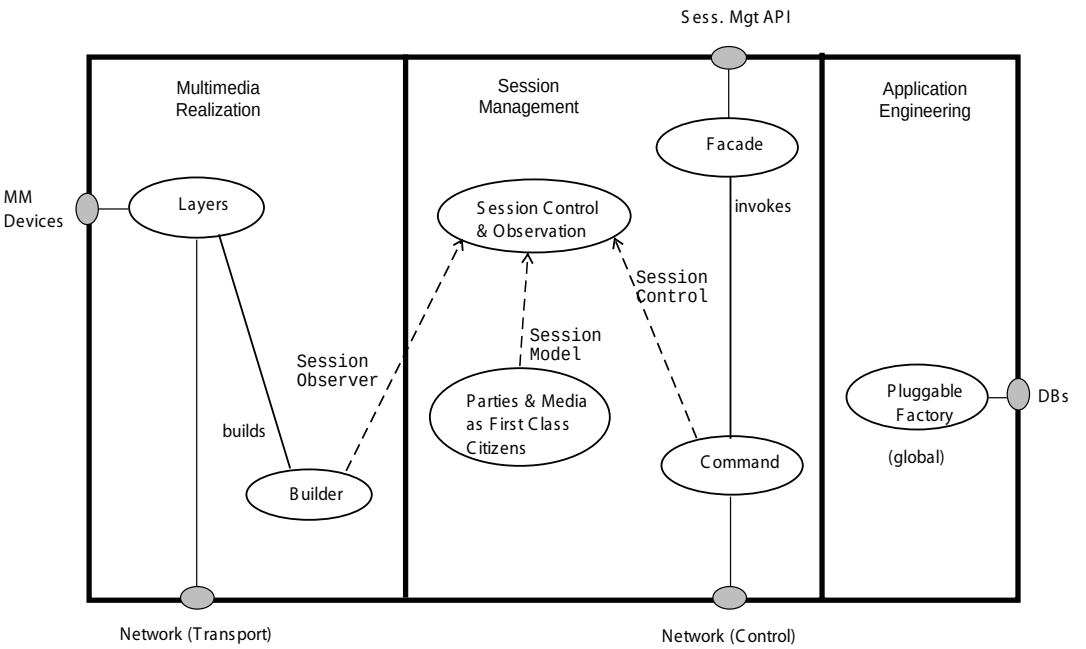
solving a specific problem like request to events, GUIs, and on-demand objects creation. In object-oriented technologies, a design pattern is represented by a specific organization of classes and relationships that may be implemented using any object-oriented language. The book by Gamma, Helm, Johnson, and Vlissides (1995) is an anchor reference for design patterns. It establishes (a) the four essential elements of a pattern, namely, the pattern name, the problem, the solution and the consequences and (b) a preliminary catalog gathering a set of general purposes patterns. Later, many application-specific software patterns have been proposed such as in multimedia, distributed environments and security.

Compared to software frameworks discussed earlier, patterns can be considered as micro software frameworks; a partial program for a problem domain. They are generally used as building blocks for larger software frameworks.

MediaBuilder (Van den Broecke & Coplien, 2001) is one of most successful initiatives to pattern-oriented architectures for mobile multimedia applications. MediaBuilder is a services platform that enables real-time multimedia communication (i.e., audio, video, and data) between end-user PC's. It supports value-added services such as multimedia conferencing, telelearning, and teleconsultation, which allows end-users at different locations to efficiently work together over long distances. The software architecture is a set of patterns combined together to support session management, application protocols, and multimedia devices. Figure 3 summarizes the main patterns brought into play in order to determine the basic behavior of MediaBuilder. Each pattern belongs to one of the functional areas, namely; multimedia realization, session management, and application engineering.

The use of design patterns for mobile multimedia is driven by the desire to provide a powerful tool for structuring, documenting, and communicating the complex software architecture. They also allow the use of a standard language making the overall

Figure 3. Architecture of MediaBuilder patterns (Van den Broecke & Coplien, 2001)



architecture of the multimedia application easier to understand, extend, and maintain.

The synergy of the three techniques previously discussed is depicted in Schmidt and Buschmann (2003). This synergy contributes to mobile multimedia development by providing high quality software architectures.

CHALLENGES OF MOBILE MULTIMEDIA SOFTWARE ENGINEERING

While system support for multimedia applications has been seriously investigated for several years now, the software engineering community has not yet reached a deep understanding of the impacts of "mobility" for multimedia systems. The latter has additional requirements compared to traditional multimedia applications. These requirements are linked to the versatility of the location of consumers and the diversity of their preferences. In the following, we address the main research areas that must be investigated by the software engineering community in supporting the development of mobile multimedia applications. These areas are not orthogonal. It means that same or similar research items and issues appear in more than one research area. We have divided the research space into four key research areas: (1) mobility, (2) context-awareness, and (3) real-time embedded multimedia systems.

Mobility

For the purpose previously discussed, the first trail to investigate is obviously "mobility." It is viewed by Roman, Picco, and Murphy (2000) to be the study of systems in which computational components may change location. In their roadmap paper on software engineering for mobility, they approach this issue from multiple views including models, algorithms, applications, and middleware. The middleware approach is generally adopted for the purpose of hiding hardware heterogeneity of mobile platforms and to provide an abstraction layer on top of specific APIs for handling multimedia content.

However, current investigations of software engineering for mobility argue that there is a lack of well-confirmed tools and techniques.

Context-Awareness

Context has been considered in different fields of computer science, including natural language processing, machine learning, computer vision, decision support, information retrieval, pervasive computing, and more recently computer security. By analogy to human reasoning, the goal behind considering context is to add adaptability and effective decision-making.

In general mobile applications, context becomes a predominant element. It is identified as any information that can be used to characterize the situation of an entity. Where an entity is a person, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves (Dey, 2001). Context is heavily used for e-services personalization according to consumers' preferences and needs and for providing fine-grained access control to these e-services. In the domain of mobile multimedia, this rule is still valid. Indeed, multimedia content whether this content is static (e.g., jpeg, txt), pre-stored (e.g., 3gp, mp4) or live, must be tuned according to the context of use. Mobile cinema (Pan, Kastner, Crowe, & Davenport, 2002) is an example, it is of great interest to health, tourism, and entertainment. Mobile cinema relies on broadband wireless networks and on spatial sensing such as GPS or infrared in order to provide mobile stories to handled devices (e.g., PDAs). Mobile stories are composed of media sequences collected from media spots placed in the physical location. These sequences are continually rearranged in order to form a whole narrative. Context used to assemble

mobile stories are mainly time and location but can be extended to include information collected using bio-sensors and history data.

Multimedia mobile service (MMS) is a brand new technology in the market but rapidly becomes a very popular technique used to exchange pictorial information with audio and text between mobile phones and different services. Häkkilä and Mäntyjärvi (2004) propose a model for the combination of location—as context—with MMS for the provision of adaptive types of MM messages. In their study, the authors explore user experiences on combining location sensitive mobile phone applications and multimedia messaging to novel type of MMS functionality. As they state in Häkkilä & Mäntyjärvi (2004), the selected message categories under investigation were presence, reminder, and notification (public and private), which were selected as they were seen to provide a representing sample of potentially useful and realistic location related messaging applications.

Coming back to the software perspective and based on a review of current context-aware applications, Ghita Kouadri Mostéfaoui (2004) points up to the lack of reusable architectures/mechanisms for managing contextual information (i.e., discovery, gathering, and modeling). She states that most of the existing architectures are built in an ad hoc manner with the sole desire to obtain a working system. As a consequence, context acquisition is highly tied up with the remaining infrastructure leading to systems that are difficult to adapt and to reuse.

It is clear that context-awareness constitute a primordial element for providing adaptive multimedia content to mobile devices. Even if currently, location is the most used source of contextual information, many other types can be included such users' preferences. Thus, we argue that leveraging mobile multimedia software is tied up with the improvement of software engineering for context-awareness. The latter constitutes one of the trails that should be considered for the development of adaptive mobile multimedia applications.

Real-Time Embedded Multimedia Systems

Real-time synchronization is an intrinsic element in multimedia systems. This ability requires handling events quickly and in some cases to respond within specified times. Real-time software design relies on specific programming languages in order to ensure that deadlines of system response are met. Ada is an example language; however, for ensuring a better performance, most real-time systems are implemented using the assembler language. The mobility of multimedia applications introduces additional issues in handling time constraints. Such issues are management of large amount of data needed for audio and video streams. In Oh and Ha (2002), the authors present a solution to this problem by relying on code synthesis techniques. Their approach relies on buffer sharing. Another issue in real-time mobile multimedia development is software reusability. Succi, Benedicenti, Uhrik, Vernazza, and Valerio (2000) point to the high importance of reusability for the rapid development of multimedia applications by reducing development time and cost. The authors argue that reuse techniques are not accepted as a systematic part of the development process, and propose a reusable library for multimedia, network-distributed software entities.

Software engineering real-time systems still present many issues to tackle. The main ones are surveyed by Kopetz (2000) who states that the most dramatic changes will be in the fields of composable architectures and systematic validation of distributed fault-tolerant real-time systems.

Software engineering mobile multimedia embraces all these domains and therefore claims for accurate merging of their respective techniques and methodologies since the early phases of the software development process.

Bridging the Gap Between Software Engineering and Mobile Multimedia

Different software engineering techniques have been adopted to cope with the complexity of designing mobile multimedia software. Selecting the "best" technique is a typical choice to be made at the early stage of the software design phase. Based on the study we presented earlier, we argue that even if the research community has been aware of the advantages of software engineering for multimedia, mobility of such applications is not yet considered at its own right. As a result, the field is still lacking a systematic approach for specifying, modeling and designing, mobile multimedia software. In the following, we stress a preliminary set of guidelines for the aim to bridging the gap between software engineering and mobile multimedia.

- The mobile multimedia software engineering challenges lie in devising notations, modeling techniques, and software artifact that realize the requirements of mobile multimedia applications including mobility, context-awareness, and real-time processing
- The software engineering research can contribute to the further development of mobile multimedia by proposing development tools that leverage the rapid design and implementation of multimedia components including voice, image, and video
- Training multimedia developers to the new software engineering techniques and methodologies allows for the rapid detection of specific tools that leverage the advance of mobile multimedia
- Finally, a community specializing in software engineering mobile multimedia should be established in order to (1) gather such efforts (e.g., design patterns for mobile multimedia) and (2) provide a concise guide for multimedia developers (3) to agree on standards for multimedia middleware, frameworks and reusable multimedia components

CONCLUSION

In this chapter, we highlighted the evolving role of software engineering for mobile multimedia development and discussed some of the opportunities open to the software engineering community in helping shape the success of the mobile multimedia industry. We argue that a systematic reliance on software engineering methodologies since the early stages of the development cycle is one of the most boosting factors of the mobile multimedia domain. Developers should be directed to use reuse techniques in order to reduce maintenance costs and produce high-quality software even if the development phase takes longer.

REFERENCES

Coffland, J. E., & Pimentel, A. D. (2003). A software framework for efficient system-level performance evaluation of embedded systems. *Proceedings of the 18th ACM Symposium on Applied Computing, Embedded Systems Track*, Melbourne, FL (pp. 666-671).

Davidyuk, O., Riekki, J., Ville-Mikko, R., & Sun, J. (2004). Context-aware middleware for mobile multimedia applications. *Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia* (pp. 213-220).

Dey, A. (2001). Supporting the construction of context-aware applications. In *Dagstuhl Seminar on Ubiquitous Computing.*

Ericsson. (2005). *EDGE introduction of high-speed data in GSM/GPRS networks,* White paper. Retrieved from http://www.ericsson.com/products/white_papers _pdf/edge_wp_technical.pdf

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design patterns: Elements of reusable object-oriented software*. Reading, MA: Addison-Wesley.

GSM World. (2005). *GPRS Platform*. Retrieved from http://www.gsmworld.com/technology/gprs/intro.shtml#1

Häkkilä, J., & Mäntyjärvi, J. (2004) User experiences on combining location sensitive mobile phone applications and multimedia messaging. *International Conference on Mobile and Ubiquitous Multimedia,* Maryland (pp. 179-186).

Hirakawa, M. (1999). Do software engineers like multimedia? *Proceedings of the International Conference on Multimedia Computing and Systems,* Florence, Italy (pp. 85-90).

Jain, R. (2001). Mobile Multimedia. *IEEE MultiMedia, 8*(3), 1.

Kopetz, H. (2000). Software engineering for real-time: A roadmap. *Proceedings of the Conference on the Future of Software Engineering.*

Koskinen, J. (2003). *Software maintenance costs.* Information Technology Research Institute, ELTIS-project, University of Jyväskylä.

Kouadri Mostéfaoui, G. (2004). *Towards a conceptual and software framework for integrating context-based security in pervasive environments.* PhD thesis. University of Fribourg and University of Pierre et Marie Curie (Paris 6), October 2004.

LeClaire, J. (2005). Demand for mobile multimedia services on rise. *E-Commerce Times.* Retrieved from http://www.ecommercetimes.com/story/Demand-for-Mobile-Multimedia-services-on-Rise-40168.html

Mohapatra, S., Cornea, R., Nikil, D., Dutt, N., Nicolau, A., & Venkatasubramanian, N. (2003). Integrated power management for video streaming to mobile handheld devices. *ACM Multimedia 2003* (pp. 582-591).

Nakajima, T. (2002). Experiences with building middleware for audio and visual networked home appliances on commodity software. *ACM Multimedia 2002* (pp. 611-620).

Nokia Inc. (2005). *Mobile entry.* Retrieved from http://www.nokia.com/nokia/0,6771,5648 3,00.html

Oh, H., & Ha, S. (2002). *Efficient code synthesis from extended dataflow graphs for multimedia applications.* Design Automation Conference.

Pan, P., Kastner, C., Crowe, D., & Davenport, G. (2002). M-studio: An authoring application for context-aware multimedia. *ACM Multimedia 2002* (pp. 351-354).

Roman, G. C., Picco, G. P., & Murphy, A. L. (2000). Software engineering for mobility: A roadmap. In A. Finkelstein (Ed.), *Future of software engineering. ICSE'00,* June (pp. 5-22).

Scherp, A., & Boll, S. (2004). *Generic support for personalized mobile multimedia tourist applications.* Technical Demonstration for the ACM Multimedia 2004, New York, October 10-16.

Schmidt, D. C., & Buschmann, F. (2003). Patterns, frameworks, and middleware: Their synergistic relationships. *Proceedings of the 25th International Conference on Software Engineering (ICSE 2003)* (pp. 694-704).

Succi, G., Benedicenti, L., Uhrik, C., Vernazza, T., & Valerio, A. (2000). Reuse libraries for real-time multimedia over the network. *ACM SIGAPP Applied Computing Review, 8*(1), 12-19.

UMTS. (2005). *UMTS*. Retrieved from http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci213688,00.html

Van den Broecke, J. A., & Coplien, J. O. (2001). Using design patterns to build a framework for multimedia networking. *Design patterns in communications software* (pp. 259-292). Cambridge University Press.

KEY TERMS

Context-Awareness: Context awareness is a term from computer science that is used for devices that have information about the circumstances under which they operate and can react accordingly.

Design Patterns: Design patterns are standard solutions to common problems in software design.

Embedded Systems: An embedded system is a special-purpose computer system, which is completely encapsulated by the device it controls.

Middleware: Middleware is software that can significantly increase reuse by providing readily usable, standard solutions to common programming tasks, such as persistent storage, (de)marshalling, message buffering and queuing, request de-multiplexing, and concurrency control.

Real-Time Systems: Hardware and software systems that are subject to constraints in time. In particular, they are systems that are subject to deadlines from event to system response.

Software Engineering: Software engineering is a well-established discipline that groups together a set of techniques and methodologies for improving software quality and structuring the development process.

Software Frameworks: Software frameworks are reusable foundations that can be used in the construction of customized applications.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 251-265, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Section 5

Organizational and Social Implications

This section includes a wide range of research pertaining to the social and organizational impact of multimedia technologies around the world. Chapters introducing this section analyze multimedia as a vehicle for cultural transmission and language, while later contributions offer an extensive analysis of educational multimedia. The inquiries and methods presented in this section offer insight into the integration of multimedia technologies in social and organizational settings while also emphasizing potential areas of study within the discipline.

Chapter 5.1
Multimedia as a Cross-Channel for Cultures and Languages

Ramesh C. Sharma
Indira Gandhi National Open University, India

Sanjaya Mishra
Indira Gandhi National Open University, India

INTRODUCTION

Around the world many communities have been constantly struggling to maintain their customs, traditions and language. Many communities have been on the move from place to place due to various factors of social change, such as war, search of food, land, and climatic calamities. Such forces have given rise to different cultures and languages through fusion or the creation of new cultures. The cultures not only exist within nationalities and ethnic groups, but also within communities, organizations and other systems. A language is an integral component of cultural identification (Rogers & Steinfatt, 1999). Matsumoto (1996, p. 16) defined culture as, "the set of attitudes, values, beliefs, and behaviours shared by a group of people, but different for each individual, communicated from one generation to the next." A culture is dynamic in nature; if static, it will cease or lose its identity in due course of time. Cultural values are affected and reinforced by languages. A language is a representation of a different way of thinking as well as a different way of speaking. Languages have significant influence on the cognition (Gudykunst & Asante, 1989; Pincas, 2001).

WHAT IS MULTIMEDIA?

Recent advances in information and communication technologies (ICTs) have resulted in the integration of the basic multimedia technology with the personal computer. Thus now it is possible to offer pedagogically useful services through this interactive medium (Peters, 2003). In its simplest form multimedia can be defined as, "an integration of multiple media elements (audio, video, graphics, text, animation etc.) into one synergetic and symbiotic whole that results in more benefits for the end user than any one of the media elements

can provide individually" (Reddi, 2003, p. 3). One of the basic advantages of multimedia tools lies in presenting learning materials in multiple (i.e. audio, visual and textual) formats. Jacobson and Spiro (1995) argued that complex information is learned more effectively if the learning experiences are presented in multimedia formats. Learners' interests and motivations can be increased by the integration of rich and dynamic multimedia into the learning experiences (Smith & Jones, 1989). Student learning can also be effectively increased by combining multi-modal dynamic simulations with audio, when the audio media is an integral part of information to be learned (Moreno & Mayer, 2000). Hoyer (1999) experienced effective teaching, research, counseling and learner support when interactive multimedia modules are integrated with services such as teleconferencing.

Multimedia has many uses in education as an instructional tool and as a product development tool. The booster factor to the development of multimedia technologies has been its pedagogical implications and effects on teaching and learning practices. Incorporating multimedia elements supports a paradigm shift in the pedagogy. The traditional teacher-centered or technology-centered approach of multimedia instruction has been supplanted by the constructivist learner-centered approach (Relan & Gillani, 1997; LeFoe, 1998; Richards & Nason, 1999; Tearle, Dillon & Davis, 1999; Abbey, 2000). Multimedia applications have created new opportunities for instructional designers to present instruction through dynamic integration of words, static and dynamic graphics and verbal information. Multimedia technologies have been found to be useful in enhancing learning and learner satisfaction, and increasing the visibility and appeal of existing programs. These technologies also support portability, modularity, visualization, efficiency in instructional design, and learning consistency (Oberlin, 1996; Hede, 2002; Yildirim, Ozden, & Aksu, 2001).

MULTIMEDIA IN CULTURAL CONTEXT

Communities and societies over time have adopted different measures for the preservation, transmission and advancement of their languages and cultural heritages. Education and communication are two effective measures. Owing to the dynamism of communities, the educational sector has been constantly witnessing structural, pedagogical, procedural and technological changes. These changes are inevitable. Increase in the demand for learning and increase in socio-cultural infusion has resulted in increasing pressure on educational providers for new ways of delivering instructional programs. The world has expanded and contracted in terms of population and space respectively. As a direct consequence, different cultures and languages have realized the increasing importance of having a dynamic and vibrant means of communication in place, which can help them maintain their identity, and observe progress in a multicultural learning environment. Odasz (n.d.) states, "The world's diverse cultures jointly represent the full cultural genome of humankind's search for individual and group identity and meaning" and exert pressure to record this important "shared story of humankind". The sooner actions are taken to save the cultural knowledge of our ancestors, the better will be, as it is feared that nearly half of the world 6000 languages may disappear in one lifetime. Odasz (n.d.) recommends, "The vast cultural knowledge of our elders must be recorded via multimedia storytelling for preservation while they (our elders) are still with us."

UNESCO has also been very concerned about the preservation of cultures of indigenous people (around 350 million individuals in more than 70 countries) representing more than 5000 languages and cultures. The cultural heritage of some of them is nearing extinction, if no action is taken to save them (http://portal.unesco.org/culture/en/ev.php-URL_ID=2946&URL_DO=DO_TOPIC&URL_

SECTION=201.html). On 10 December 1994, the United Nations General Assembly declared the *International Decade of the World's Indigenous People*. To mark the significance of cultural diversity of indigenous populations, the *International Day of the World's Indigenous Peoples* is celebrated every year on 9 August.

The transfer of knowledge from one region to another or from one generation to another requires educational or learning material understandable to many. A Universal Networking Language (UNL) is being developed by the United Nations University. The UNL is the global lingua franca of computers and will be providing services for any pair of languages from over 180 countries. Yoshii, Katada, Alsadeqi, and Zhang (2003) found that although the content on the Internet is growing, much of it is written in English, rendering it unsuitable for those who do not understand English. If they wish to access that content they are made to learn English, and thus for such non-English speaking people, the language learnt sometimes is inappropriate. Tools have been developed to translate material from one language to other, but such translation may not be free from errors and may misrepresent the actual sense of context, and may also suffer from cultural imperialism (Phillipson, 2002). Thus it becomes very pertinent to keep certain issues related to cultural sensitivity and languages in mind while developing multimedia content.

CULTURAL SENSITIVITY FOR MULTIMEDIA DESIGNING

Pruitt-Mentle (2003) revealed that people with different cultural backgrounds have different expectations and attitudes towards educational software. And educational software developed in one country may suffer from cultural biases. To overcome cross-cultural barriers of verbal and non-verbal communication or symbols as practiced in a particular country, Pruitt-Mentle recommended the learning of such cultural differences while designing multimedia learning resources, rather than to deny or ignore such diversity.

Since a variety of personnel are involved in instructional design, Gunawardena, Wilson and Nolla (2003) advocated understanding of the cultures of all involved, such as course designers, teachers and students. Everyone has his own background culture and attitudes and societal norms on the instructional product. An essential concern while designing multimedia applications for diverse community groups is to be aware of cultural influences. This creates the need to develop cultural sensitivity and awareness among designers and instructors (Powell, 1997; Reeves, 1997). Chen, Mashhadi, Ang and Harkrider (1999) suggested, "the interface designer must be aware how different cultures will respond to issues of the layout of the graphical interface, image, symbols, colour and sound, impressing that culture itself cannot be objectified unproblematically as just another factor to be programmed into a learning course" (p. 220). The multimedia designers, while designing for a diverse community, must be flexible about a variety of presentation formats to accommodate cross-cultural differences (Collis, 1999).

Other issues which may arise while developing culturally sensitive instructional material are values and attitudes, ethical perspectives diversity within groups, historical perspectives, socio-economic perspectives, social roles, social networks, learner expectations, learning styles, and opportunities for interactions (Dunn & Griggs, 1995; Powell, 1997a, 1997b; Collis, 1999; Mcloughlin, 1999; Reeves, 1997). Gjedde (2005) suggested that issues such as social and cultural backgrounds and gender-specific interests, must be considered when developing meaningful content and creating narrative multimedia learning environments.

MEETING THE CHALLENGE OF CULTURES AND LANGUAGES THROUGH MULTIMEDIA

Mouafo and Miller (2002) found the use of digital and web-based multimedia cartography very useful in injecting new life into archived data such as old paper maps and photographs. Mouafo and Miller presented the results of a "The Historical Evolution of Iqaluit" mapping project showing the changes in time and space in the city over the past 50 years (1948-1998). (Iqaluit, is the upcoming capital city of the new territory of Nunavut in Northern Canada). Such projects have high significance to the culture, education, and tourism as well as enhancing city planning and development. Multimedia maps have many advantages—they enable a dynamic and multifaceted representation of space and time, superior map production and dissemination, improved information and knowledge transfer, and greater map accessibility (Balram & Dragicevic, 2005).

UCLA has undertaken the development of Multimedia Research Projects in areas like Anthropology, Chemistry/Biochemistry, Computer Science, Dance, Dentistry, Education, Electrical Engineering, Germanic Languages, Linguistics, Medicine, Music Composition, Musicology/Performance, Nursing, Physics, Psychology, Public Policy, Slavic Languages and Literatures, Theater, Film and Television, University Elementary School, University Extension, and Visual Library. (See details at http://www.research.ucla.edu/media.htm). These projects cover a vast array of human knowledge and activity. Some projects related to cultures and languages are:

- *Culture and Communication* (http://www.research.ucla.edu/mmedia/anthro.htm#anthro2),
- *Educational Multimedia on Pacific Studies* (http://www.research.ucla.edu/mmedia/pacrim.htm#anchor435555),
- *Crosslinguistic Research Project* (http://www.humnet.ucla.edu/humnet/al/clrl/crphome.html) for Applied Linguistics,
- *Bulgarian Traditional Music on CD-ROM* (http://www.research.ucla.edu/mmedia/ethno.htm)
- *World Music Navigator* (http://www.research.ucla.edu/mmedia/ethno.htm#anchor6071659) developing under Ethnomusicology,
- *Comparative Court Cultures In Cross-Cultural Perspective* (http://www.research.ucla. edu/mmedia/cmrs.htm) developed by The Center for Medieval and Renaissance Studies.

The UCLA School of Dentistry has developed a unique "*Preceptorship (Specialty Training) Program*" (http://www.dent.ucla.edu/ce/in-person/precep/index.html) targeted at foreign-educated dentists all over the world. This program is delivered via the Internet and aims to enrich the educational experiences of dentists in a certain specialty.

Multimedia applications have also addressed social causes and concerns. One of these projects is a multimedia AIDS prevention project "AIDS: Interactive Situations" that has been undertaken jointly by research teams from the co-operative group, CLAPS, of Pordenone, Italy, and the University of Barcelona (ICE), Spain (Illera, 2005). The cultural and linguistic differences affect the education of differently abled students in classroom. Allen (1994) and Schildroth and Hotto (1996) studied the problems of deaf children from Spanish-speaking homes in the US towards learning English and were found to perform at lower levels from their Caucasian deaf peers. If such students come from different cultures or religion, they may use different sign languages. Such a situation may create problems for students learning subjects in a particular language. To promote the Mexican-American heritage of such deaf students and to accommodate their cultural and language needs, multimedia labs have been

established and multimedia stories on CD-ROMs have been produced (Pollar, 1993; Hernandez, 1994).

MULTIMEDIA IN A FOREIGN LANGUAGE INSTRUCTIONAL CONTEXT

Many cultures (like rural folk in Rajasthan State in India) use narration and story telling as a way of passing on the past to the present generation. Narration enables the learners to participate in the process and becomes the core of an experiential and contextual approach to learning. A project in the Danish Language and Art curriculum, called, "The Narrative Universes of Children" and "Narrative in interactive Web-based learning environments" enables children to identify with the characters and situations in the narratives (Gjedde, 2005). Multimedia materials may even provide the opportunity to mimic the patterns of traditional acquisition of native language and cultural competence, which in many cultures have formerly relied on repeated observation, playful pursuit and accompanying (more or less informal) explanation instead of abstract, decontextualized texts (Dürr, 1998).

LeLoup and Ponterio (1998) listed some of the multimedia applications (e.g., audio, digitized video) on the Internet using multimedia in a Foreign Language instructional context. Some of these applications are used for less- commonly-taught (LCT) languages and offer to the viewer or learner expression of native speaker pronunciation, cultural topics and references, and techniques of mastering difficult non-Roman scripts. Here is a listing of some of these sites:

- http://www.cortland.edu/flteach/civ/ deals with French instruction;
- http://web.uvic.ca/german/149/ offers lessons for beginning German students;
- http://carla.acad.umn.edu/vpa/hebrew/intro.html provides interactive lessons in Hebrew;
- http://carla.acad.umn.edu/vpa/India/exercises.html has applications in basic Hindi language courses;
- http://www.stanford.edu/~muratha/ is useful for Swahili classes and narrates cultural stories;
- http://www.fix.co.jp/kabuki/kabuki-j.html offers many vocal and musical sounds of traditional Kabuki; and
- http://www.wavefront.com/~swithee/dictionary/welcome.html contains Russian dictionary with sounds and still images.

Multimedia applications have also been developed for a flexible, effective and efficient mode of delivery for learning and teaching Chinese, especially Chinese characters. These applications have attracted a good response, acknowledgement and better inclination of students towards multimedia mode of delivery than traditional mode of learning Chinese characters. The difficulties in learning Chinese characters can be easily tackled through multimedia; Wang (1999) found that multimedia raised learners' motivation and self-esteem, satisfying individual differences, abilities and learning styles.

CONCLUSION

The swift developments in multimedia technology and products have revolutionized the ways of teaching and learning across different cultures. The content generated with multimedia applications act as a vehicle for knowledge that is well-suited to many cultures and languages across the globe. The new multimedia tools provide great opportunities for the preservation, expression, strengthening and dissemination of cultural identities and languages. The various projects, products, tools and Internet sites discussed above truly reflect the unparalleled capabilities of multi-

media technologies towards acquiring, mastering and advancing the past heritage to different platforms. These tools can be used as self-study aids, training tools, traditional narrative resources, and social and interpersonal communication channels. Research has shown that learning becomes a more effective and pleasant process through the use of interactive multimedia. Multimedia has excellent scope in educational, health, agriculture, social, scientific, engineering, history and geographical areas. Multiple representation, multilingualism and multimodal dialogues infuse interactivity, accessibility, understanding, and knowledge extraction and summarization to the cause of many cultures and languages.

REFERENCES

Abbey, B. (Ed.). (2000). *Instructional and cognitive impacts of web-based education.* Hershey, PA: Information Science Publishing.

Ainsworth, S., & Van Labeke, N. (2002). Using a multi-repersentational design framework to develop and evaluate a dynamic simulation environment. In R. Ploetzner (Ed.), *International workshop on dynamic visualizations and learning.* Tubingen, Germany: Knowledge Media Research Centre.

Allen, T. (1994). *Who are the deaf and hard-of-hearing students leaving high school and entering postsecondary education?* Washington, DC: Office of Special Education and Rehabilitative Services, U.S. Department of Education.

Balram, S. & Dragicevic, S. (2005). An embedded collaborative systems model for implementing ICT-based multimedia cartography teaching and learning. In S. Mishra & R.C. Sharma (Eds.), *Interactive multimedia in education and training* (pp. 306-326). Hershey, PA: Idea Group Publishing.

Bodemer, D., & Ploetzner, R. (2002). Encouraging the active integration of information during learning with multiple and interactive representations. In R. Ploetzner (Ed.), *International workshop on dynamic visualizations and learning.* Tubingen, Germany: Knowledge Media Research Centre.

Bork, A. et al. (1992). The Irvine-Geneva course development system. *Proceedings of IFIP, Madrid*, 253-261.

Chen, A.Y., Mashhadi, A., Ang, D., & Harkrider, N. (1999). Cultural issues in the design of technology-enhanced learning systems. *British Journal of Educational Technology*, *30*(3), 217-230.

Collis, B. (1999). Designing for differences: cultural issues in the design of the www-based course-support sites. *British journal of Educational Technology*, *30*(3), 201-215.

Dunn, R. & Griggs, S.A. (1995). *Multiculturalism and learning style: Teaching and counseling adolescent.* Westport, CT: Praeger.

Dürr, Michael. (1998). Multimedia materials for native language programs. In Erich Kasten (Ed.), *Bicultural education in the North—Ways of preserving and enhancing indigenous peoples' languages and traditional knowledge* (pp. 269-274). Münster, New York: Waxmann Verlag GmbH.

Elsom-Cook, M. (2001). *Principles of interactive multimedia*. London: McGraw Hill.

Evans, J. (2002). *The FILTER generic image dataset: A model for the creation of image-based learning & teaching resources.* Paper presented at the ASCILITE 2002, 19th annual conference of the Australasian Society for Computers in Learning in Tertiary Education, Auckland, NZ.

Gjedde, L. (2005). Designing for learning in narrative multimedia environments. In S. Mishra & R.C. Sharma (Eds.), *Interactive multimedia in education and training* (pp. 101-111). Hershey, PA: Idea Group Publishing.

Gudmundsdottir, S. (1991). Story-maker, story-teller: narrative structures in curriculum. *Journal of Curriculum Studies, 23*(3), 207-218.

Gudykunst, W.B. & Asante, M. (1989). *Handbook of international and intercultural communication*. Newbury Park, CA: SAGE.

Gunawardena, C.N., Wilson, P.L., & Nollo, A.C. (2003). Culture and online education. In M.G. Moore & W.G. Anderson (Eds.), *Handbook of distance education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, E.T. (1998). The power of hidden differences. In M.J. Bennett (Ed.), *Basic concepts of intercultural communication: selected readings* (pp. 53-67). Yarmouth, ME: Intercultural Press.

Hede, A. (2002). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia, 11*(2), 177-191.

Hernandez, Monica (1994). Multimedia for Hispanic deaf. In Bob Hoffman (Ed.), *The encyclopedia of educational technology.* San Diego State University. Retrieved October 24, 2004 from *http://coe.sdsu.edu/eet/articles/multitech/*

Hoyer, H. (1999). Lernvaum irtuelle Universitat: Challenge and opportunity for the fernuniversitat. In G.E. Ortner & F. Nickolmann (Eds.), *Socio-economics of virtual universities* (pp. 213-222). Weinheim, Germany: Beltz.

Illera, J. L. R. (2005). Interactive multimedia and AIDS prevention: A case study. In S. Mishra & R.C. Sharma (Eds.), *Interactive multimedia in education and training* (pp. 271-288). Hershey, PA: Idea Group Publishing.

Jacobson, M.J. & Spiro, R.J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research, 12*, 301-333.

Kinshuk, & Patel, A. (2003). Optimizing domain respresentation with multimedia objects. In S. Naidu (Ed.), *Learning and teaching with technology: principles and practice* (pp. 55-68). London and Sterling, VA: Kogan Page.

Laurillard, D., Stratfold, M., Luckin, R., Plowman, L., & Taylor, J. (1998). Multimedia and the learner's experience of narrative. *Computers and Education*, *31*, 229-242.

Lefoe, G. (1998). Creating Constructivist Learning Environments on the Web: The Challenge of Higher Education. *Paper presented at ASCILITE '98 Conference*, Wollongong. Retreived March 20, 2003, from *http://www.ascilite.org.au/conferences/wollongong98/ascpapers98.html*

LeLoup, Jean W. & Ponterio, R. (1998). Using WWW multimedia in the foreign Language classroom: Is this for me? *Language Learning & Technology*, *2*(1), 4-10. Online: *http://llt.msu.edu/vol2num1/Onthenet/*

Matsumoto, D. (1996). *Culture and psychology.* Pacific Grove, CA: Brooks/Cole.

McLoughlin, C. (1999). Culturally responsive technology use: Developing an on-line community of learners. *British Journal of Educational Technology, 30*(3), 231-243.

Moreno, R. & Mayer, R.E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*, 117-125.

Mouafo, D. & Muller, A. (2002). *Web based multimedia cartography applied to the historical evolution of Iqaluit, Nunavat.* Paper presented at Symposium on Geospatial Theory, Processing and Applications, Ottawa. Retrieved October 14, 2004 from *http://www.isprs.org/commission4/proceedings/pdfpapers/501.pdf*

Oberlin, J. L. (1996). The financial mythology of information technology: The new economics. *Cause/Effect*, 21-29.

Odasz, Frank (n.d.) *Realizing cultural sovereignty through Internet applications*. Retrieved October 14, 2004, from *http://lone-eagles.com/sovereignty. htm*

Peters, O. (2003). Learning with new media in Distance Education. In Michael Grahame Moore and William G. Anderson (Eds.), *Handbook of distance education* (p. 88). Mahwah, NJ: Lawrence Erlbaum Associates.

Phillipson, R. (2002). Global English and local language policies. In A. Kirkpatrick (Ed.), *Englishes in Asia: Communication, identity, power and education*. Melbourne: Language Australian Ltd.

Pincas, A. (2001). Culture, cognition and communication in global education. *Distance Education, 22*(1), 30-51.

Pollar, G. (1993). *Making accessible to the deaf CD-ROM reading software*. Austin, TX: Texas School for the Deaf.

Powell, G.C. (1997a, March-April). Diversity and educational technology: Introduction to special issue. *Educational Technology, 37*(2), 5.

Powell, G.C. (1997b, March/April). On being culturally sensitive instructional designer and educator. *Educational Technology, 37*(2), 6-14.

Pruitt-Mentle, D. (2003, June). *Cultural dimensions of multimedia: Design for instruction*. Presented at NECC conference, Seattle. Retrieved October 7, 2004, from *http://edtechoutreach. umd. edu/Presentations/NECC2003/2003_ presentation_v2.ppt*

Reddi, U. V. (2003). Multimedia as an educational tool. In U.V. Reddi & S. Mishra (Eds.), *Educational multimedia: A handbook for teacher-developers* (pp. 3-7). New Delhi: CEMCA.

Reeves, T.C. (1997, March-April). An evaluator looks at cultural diversity. *Educational Technology, 37*(2), 27-31.

Relan, A., & Gillani, B. (1997). Web-based instruction and the traditional classroom: Similarities and differences. In B.H. Khan (Ed.), *Web-based instruction*. NJ: Educational Technology Publications.

Richards, C., & Nason, R. (1999, March). *Prerequisite principles for integrating (not just tacking-on) new technologies in the curricula of tertiary education large classes*. Paper presented at the ASCILITE '99 Conference, Brisbane. Online: *http://www.ascilite.org.au/conferences/brisbane99/papers/papers.htm*

Rogers, E.M. & Steinfatt, T.M. (1999). *Intercultural communication*. Prospect Height. 16: Waveland Press.

Schildroth, A., & Hotto, S. (1996). Changes in student and program characteristics. *American Annals of the Deaf, 141*(2), 68-71.

Smith, S.G. & Jones, L.L. (1989). Images, imagination and chemical reality. *Journal of Chemical Education, 60*, 8-11.

Tearle, P., Dillon, P., & Davis, N. (1999). Use of information technology by English university teachers: Developments and trends at the time of the National Inquiry into higher education. *Journal of Further and Higher Education, 23*(1), 5-15.

Wang, Y. (1999). Learning Chinese Characters through Multimedia. *CALL-EJ Online, 1*(1). Retrieved October 24, 2004, from *http://www.clec. ritsumei.ac.jp/english/callejonline/4-1/wang2. html*

Wooldridge, B. (2000, July). *Foreigner talk: An important element in cross-cultural management education and training.* Paper presented at the Annual Conference of the International Associa-

tion of Schools and Institutes of Administration, Beijing, China.

Yildirim, Z., Ozden, M. Y., & Aksu, M. (2001). Comparison of hypermedia learning and traditional instruction on knowledge acquisition and retention. *The Journal of Educational Research, 94*(4), 207-214.

Yoshii, R., Katada, F., Alsadeqi, F., & Zhang, F. (2003). Reaching students of many languages and cultures. *Proceedings of the EDMEDIA Conference*, AACE.

KEY TERMS

Culture: The ideals, values, symbols and behaviors of human societies that create a distinctive identification.

Language: A means of verbal and non-verbal communication of thoughts and ways of speaking.

Multimedia: Integration of several media, such as text, audio, video, animation, etc.

This work was previously published in Encyclopedia of Distance Learning, Vol. 3, edited by C. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers, and G.A. Berg, pp. 1310-1316, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.2
Developing Culturally Inclusive Educational Multimedia in the South Pacific

Christopher Robbins
Rhode Island School of Design, USA
The University of the South Pacific, Fiji

ABSTRACT

This chapter explores how educational technology can be developed according to indigenous learning approaches of the South Pacific. It is based on an expansive research and development project conducted 2003-2004 at The University of the South Pacific (USP). After an introduction to several aspects of indigenous South Pacific learning approaches and their usage in the formal learning sector, I make several recommendations for instructional technology design based on these principles, illustrated with examples of educational technology projects that apply these recommendations. Specifically, we follow educational multimedia efforts at USP that enable learning in wholes, encourage observation and imitation and utilize vernacular metaphors and languages. This includes recommendations for interface design, interaction design and decentralized content localization.

INTRODUCTION

Information technology plays a vital but contentious role in tertiary education in the South Pacific. Although many students must rely on information technology for higher education, there are concerns regarding the intrinsic cultural biases of, and unbalanced access to, educational technology in the region (Matthewson, 1994; Thaman, 1997; Va'a, 1997; Wah, 1997). Approximately half of the 15,000 students of The University of the South Pacific (USP) study through USP's Distance and Flexible Learning (DFL) Centers, a network of mini-campuses in 12 island-nations (Cook Islands, Fiji, Kiribati, Marshall Islands, Nauru, Niue, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu and Vanuatu) linked through VSAT satellite (USP DFL Unit, 2004). These students negotiate audio and video conferences, Web-based group activities, interactive CD-ROMs, video broadcasts of lectures, email, faxes and even CB

radio as they communicate with their teachers and fellow students.

While educational technology offers distance students a higher degree of interaction with their educational materials, lecturers and fellow students, it can also introduce additional cultural biases into their already imported education system. The cultural gaps between USP's formal education system and the diverse South Pacific cultures of USP's staff and students has been widely documented (Lockwood, Roberts, & Williams, 1998; Thaman, 1997; Va'a, 1997, 2000; Wah, 1997). Closing these gaps through culturally inclusive curricula and pedagogy has become an institutional priority. Recently, the University's focus on culturally relevant pedagogy has broadened to include the instructional design of educational technology.

As part of this initiative, The University of the South Pacific Media Centre, with funding from the Japan International Cooperation Agency (JICA), completed a project that examined how educational multimedia can be designed according to the learning approaches of the South Pacific. In the study (Robbins, 2004), we captured the views of Pacific educationists through a series of interviews and a review of academic literature covering indigenous pedagogy in the South Pacific. We conducted interviews, employed questionnaires and usability tests with staff and students from the 12 member-nations of USP to find applications of regional learning approaches to the development of educational technology and built an educational CD-ROM to audit and illustrate the findings.

In this chapter I outline several recommendations and applications of these findings, focusing on how educational multimedia can be made culturally relevant to the South Pacific.

The goal of this chapter is not only to enable technological fluency by helping developers create educational multimedia designed specifically for the region, but also to ensure that the technology promotes indigenous approaches and values, rather than submerging them under dominant technological hegemonies.

PEDAGOGIC FOCUSES FOR EDUCATIONAL TECHNOLOGY DEVELOPERS IN THE SOUTH PACIFIC

This chapter concentrates on three regional pedagogic focuses relevant to educational technology in the region:

1. Enable learning in wholes (Thaman, 1992; Yorston, 2002);
2. Encourage observation and imitation (Lima, 2003; Pagram, Fetherston, & Rabbitt, 2000; Taufe'ulungaki, 2003; Thaman, 1999; Yorston, 2002); and
3. Utilize vernacular metaphors and languages (Afamasaga, 2002; Pagram, Fetherston, & Rabbitt, 2000; Taafaki, 2001; Taufe'ulungaki, 2003; UNESCO, 1992).

Enabling learning in wholes, or preserving "the big in the small," arises time and again as a key component of South Pacific learning approaches (Mel, 2001; Harris, 1992; Thaman, 1992; Yorston, 2002). The key concept is that rather than segmenting learning activities into distinct conceptual units, ideas are approached as they can be applied within the context of larger tasks (Thaman, 1992; Yorston, 2002). In other words, rather than master each step consecutively, learners witness and then imitate the whole, attaining the desired goal through trial and error (Mel, 2001). To preserve the whole, complex activities are tackled as "successive approximations" of the final product rather than as "sequenced parts" (Harris, 1992, p. 38). For example, in learning a musical piece using this method, a band would play the entire piece through until they had mastered it, as opposed to repeating individual refrains.

A nuance of learning in wholes is that in each step, the focus is on the specific context of the task or idea and its relevance to the broader goal, as opposed to working from generalizations towards instances of universals (Thaman, 2003; Va'a, 1997).

In educational technology, this approach has ramifications for information architecture and interface design, suggesting navigation processes that are more task than concept-focused, and content-display that continues to present the wider context while students explore details.

Closely related to the concept of learning in wholes is the process of observation and imitation, as both approaches originate from task-based learning exercises applied in "real world" settings. However, while learning in wholes translates well from traditional contexts to formal education settings, observation and imitation can manifest itself less desirably as rote memorization and surface learning (Landbeck & Mugler, 1994). Tying observation and imitation activities to application rather than memorization can encourage deep learning (Landbeck & Mugler, 1994). In other words, showing how a concept can be applied to something the student already knows can situate the knowledge more deeply than having students memorize sets of tasks or terms.

Teaching from students' own cultural contexts can be difficult with a population as diverse and distributed as that served by USP, and is further confounded by the colonial history and post-colonial legacy of formal education in the region (Petaia, 1981; Thaman, 2000b, 2003; Wah, 1997). Most textbooks used at USP are published in Europe and North America, and utilize examples from the cultures in which the books are produced (Thaman, 2000a). Despite efforts to create culturally relevant learning materials at USP, it can be incredibly difficult to choose examples and metaphors that make sense in all 12 countries served by the University:

> *Some of the course writers only use examples from the countries they know. If you look at sourcebooks, most use examples from Fiji and Samoa — a staff-member at the USP Solomon Island DFL Centre.* (Robbins, 2004, p. 29)

> *The exam paper had to do with kava. It was like double-dutch to us… Most of the examples are very Fijian. We don't have veggie markets. We don't have military management. I have to pick something we can identify with — a staff-member at the USP Nauru DFL Centre.* (Robbins, 2004, p. 29)

In addition to local metaphors, there is also a need for local languages in educational materials. Vernacular languages play a vital role in this officially English-speaking University. As a lecturer at the Samoa Campus confessed, "after class, when there are no other English-speaking students around, they ask me to explain in Samoan" (Robbins, 2004, p. 16). A program assistant's comments at the Tuvalu DFL Centre clarified, "all students understand English, it's just when it comes to tutoring, or when a point needs to be understood subtly, the students and the tutors prefer to exchange in the local language" (Robbins, 2004, p. 16).

Regional research has shown that using vernacular language aids comprehension (Pagram, Fetherston, & Rabbitt, 2000; Taafaki, 2001; Taufe'ulungaki, 2003) and deeper learning (Kalolo, 2002) where English alone can fail, and is also important for purposes of cultural preservation (Afamasaga, 2002; Sami, 2004; Taufe'ulungaki, 2003; Veramu, 2004).

Taken together, these pedagogic foci indicate that educational technology in the South Pacific should utilize vernacular metaphors, examples and languages, and should be presented in a way that enables learning in wholes using activities that rely on observation and imitation.

APPLICATIONS OF INDIGENOUS LEARNING APPROACHES TO EDUCATIONAL TECHNOLOGY IN THE SOUTH PACIFIC

In this section, I discuss how the aspects of South Pacific learning approaches outlined earlier can be applied to the development of culturally inclusive educational multimedia. Specifically, I explore how interventions at the following design levels can help achieve our pedagogic goals:

1. **Interface level:** enable learning in wholes;
2. **Interaction level:** encourage observation and imitation; and
3. **Content level:** utilize vernacular metaphors and languages.

Interface Level: Enable Learning in Wholes

The solutions we have created to enable learning in wholes are primarily interface-based, maintaining the big picture while focusing students on specifics. I discovered one very simple way to preserve "the big in the small" while looking for something else entirely. As part of our 2003-2004 study (Robbins, 2004), we asked 155 students of USP to choose between two Web designs (Figure 1). The goal was to determine whether the students would feel more comfortable with contextual, inline navigation (links within the body of the text), or with a separate menu listing the links apart from the body. I expected students to prefer the simpler, inline approach (labeled "L"), as this is closer to the layout of books, and so would be more familiar to them than the Web-derived navigation menu. However, my hunch was proven unequivocally wrong: 93% of the students preferred the menu navigation, labelled "O" (X^2 (2, N = 155) = 250.90 $p < 0.0001$). Students appreciated that the separate menu neatly summarized the longer text, and allowed them to jump directly to points of interest without losing their place.

A graphical corollary involves presenting concepts through "layers of simplicity" that offer increasing levels of detail to the students without muddying the overall purpose of the graphics. In other words, the basic lesson or overall theme of the image is always clear, even when students have drilled into details of the graphic or animation. For example, in Figure 2 we present the student with a simple initial layer of information (arrows showing immigration patterns), augmented with deeper layers of information (descriptions of archaeological remnants at different sites or related stories from regional oral histories) that the student can access by clicking sections of the map/timeline with his or her mouse.

Interaction Level: Encourage Observation and Imitation

While preserving the whole plays a role in determining the organization of content in the interface, observation and imitation have applications when designing the deeper interactivity and functionality of an educational technology project. Designing these interactions around observation and imitation enables students to act on instructions within the learning interface. For instance, in Figure 3, instructions are overlaid on the active interface, rather than being provided in separate instruction screens, so that students seeking help are shown exactly what to do, where to click and how to do it, rather than merely being provided with instructions.

By programming timers that automatically demonstrate the options for the user after a period of inactivity, this same approach can be used to help students with a tendency to "freeze" when confused, and has many applications to lab simulations.

Figure 1. Navigation preference (Robbins, 2004, used with permission)

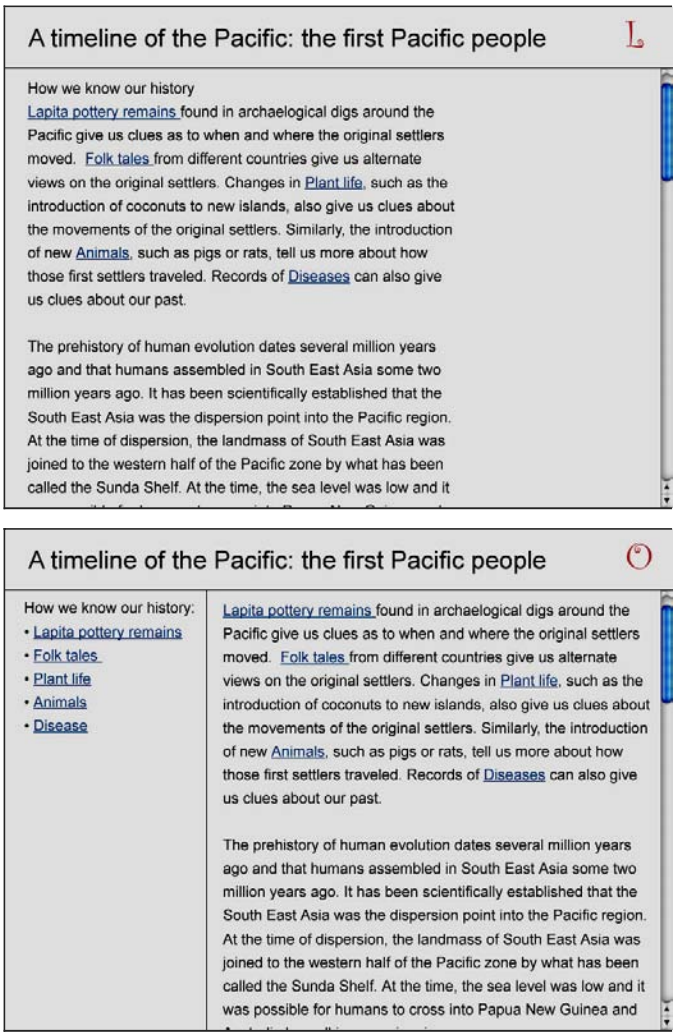
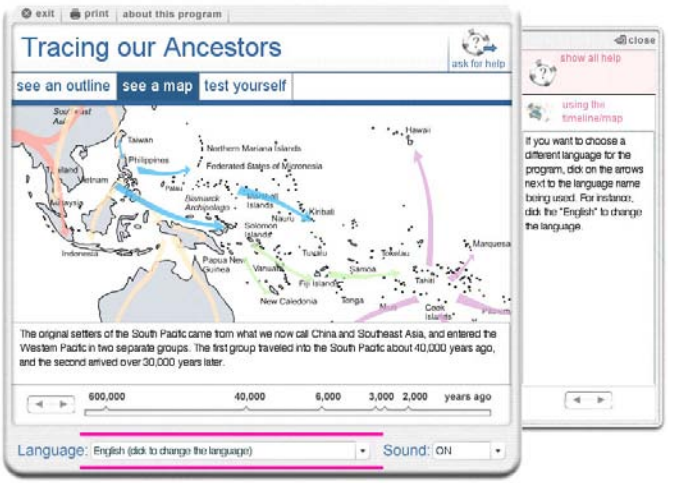


Figure 2. Using layers of simplicity on a map/timeline to show immigration patterns into the South Pacific (Robbins, 2004, used with permission)

Figure 3. Instructions are paired with an overlay showing how to act on the instructions within the interface itself (Robbins, 2004, used with permission)



Content Level: Utilize Vernacular Metaphors and Languages

The most obvious way that educational technology can be made culturally inclusive is by including the cultures of students and teachers in the educational content. This is true anywhere in the world, but is a particular focus at USP because of the wide gaps between the cultures of the educational institution, students, teachers and educational technology developers.

As USP caters to 12 distinct island-nations, providing truly local illustrations for every concept is incredibly difficult. However, by designing educational multimedia so that teachers and students can provide cultural context themselves, educational technology developers can create flexible toolsets that decentralize the contextualization process.

One such solution is the virtual peer, which uses regional examples as "seed-questions" to encourage students to contribute their own explanations. In Figure 4, peers representing different cultures in the South Pacific discuss aspects of the Internet using metaphors from their home countries. Clicking the "show me" button brings up an animation, illustration or audio clip, so the interface is not dominated by text.

To further situate the learning to the student's own circumstances, the student is asked to make his or her own descriptions of the concept (Figure 5). These answers are saved on the computer for future iterations of the program, allowing other students to view each other's perspectives.

Another approach to creating flexible educational multimedia relies on the file structure rather than the interface. For instance, in order to create multilingual and multicultural educational multimedia at USP, we have employed an open-source, three-tier file structure to help staff and students customize their learning materials themselves. Separating the core multimedia from all supporting image, text and audio files makes these updates easier to achieve. Figure 6 shows how this approach can be used to create an open framework for multilingual multimedia. Each language is given its own text file so that students and teachers need only make edits to existing text files to add additional languages and metaphors. The language files are inserted into the educational multimedia at runtime so that the materials are flexible and current.

Figure 4. A virtual peer gives his or her own descriptions of the concepts in local terms (Robbins, 2004, used with permission)

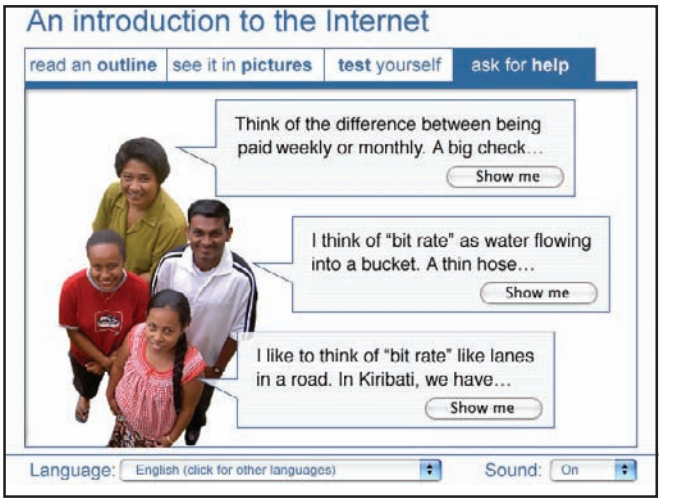
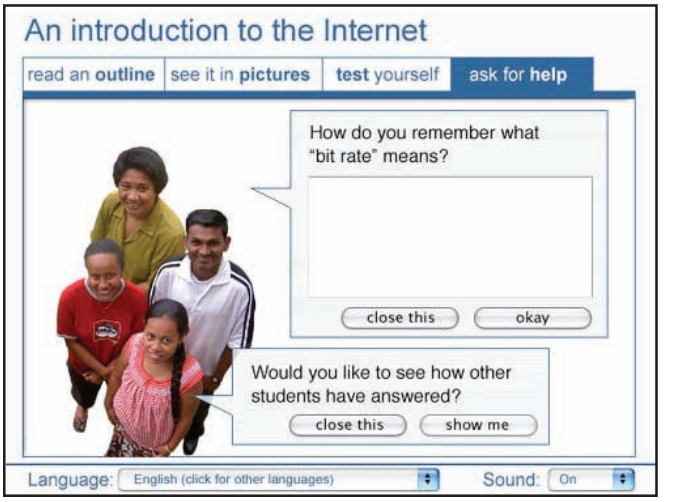


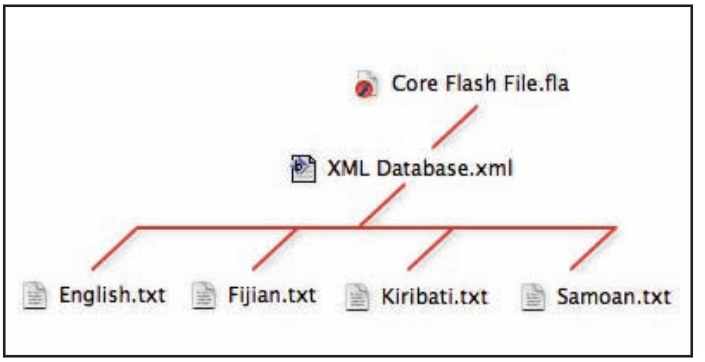
Figure 5. The student is asked to make his or her own metaphors to describe streaming media (Robbins, 2004, used with permission)



An additional benefit of the three-tier file structure is that presenting the images, audio files, videos and animations separately encourages their re-use in other media, aiding accessibility as well as customization of the media. The approach can also make the separate elements of a multimedia project (text, images, video and audio) available for re-use in other projects and media, along the lines of learning objects in object-oriented instructional design (Ahern & Van Cleave, 2002; Parrish, 2004; Wiley, 2000).

Additionally, translating entails more than simply changing the language; it involves translating the stories, devices, examples and metaphors to correlates that are meaningful within the culture of the language of translation. As such, developing educational technology that can be easily localized enables deeper reflection rather than

Figure 6. A three-tier file structure for multilingual multimedia (Robbins, 2004, used with permission)



surface memorization, encouraging students to think about their learning materials within their own cultural contexts while simultaneously creating culturally relevant educational media for future students.

While the approaches I have mentioned are tied to specific courseware, other methods for providing cultural context focus more broadly on cross-curricula databases of educational content from a variety of cultures. USP's multimedia database (http://mdb.usp.ac.fj) and cultural curriculum development initiative are the beginnings of multicultural learning toolsets for the South Pacific, with the promise of freely accessible educational media from various cultures of the region.

CONCLUSION

In order to help ensure that educational technology is built according to the cultures of the South Pacific rather than relying on de facto standards of the technology developers, the following suggestions are useful to consider:

- Offer opportunities for contextualization of the educational media, utilizing decentralized methods that enable "on-the-fly" staff and student input. Such approaches can be front-end, such as virtual peers from several countries who present examples and explain concepts using terms from their own backgrounds, or back-end, involving open file-structures that enable localization of supporting media files.
- Ensure that materials preserve the whole while offering specific anchors. For instance, long text should be accompanied by quick summaries that link to different parts of the main text. Graphical interfaces should present concepts through "layers of simplicity," in which details are available to the students without confusing the overall purpose of the graphics or the interface.
- Utilise modeling rather than separate instructions, enabling students to act on any instructions or practice any skills within the learning interface. For instance, overlay instructions on the active interface rather than providing separate instruction screens for help sections or lab simulations.
- Provide vernacular translations or glossaries within the educational multimedia. For ease of translation, save language files as separate text documents so that translators can make edits to the multimedia using simple word-processors, and need not know how to use multimedia development software.

This list is by no means exhaustive. Our own USP study (Robbins, 2004) explored how educa-

tional multimedia can help students with regard to computer access, reticence with authority figures, group learning, content-display and usability. Other studies have explored the role of terminology, multiple-modes and representational form (McCloughlin, 1999); icons and visual metaphors (Evers, et al., 1999); design and content-flow (Hedberg & Brown, 2002) and institutional/industrial collaboration (Kisum, 2003) in developing culturally inclusive educational technology.

The aspects of indigenous approaches to educational technology covered in this chapter were chosen because they reflect a broad scope of the educational technology development process (interface, interaction and content), and because their immediate applicability to a wide variety of educational technology can promote their further development in the educational technology community. Likewise, this chapter does not focus on differences between individual nations in the region, as such distinctions were beyond the scope and sample size of the original study, but rather on shared approaches that can be applied throughout the region.

REFERENCES

Afamasaga, T. (2002). Personal reflection on education in Samoa. In F. Pene, A. Taufe'ulungaki, & C. Benson (Eds.), *Tree of opportunity: Rethinking Pacific education* (pp. 96-101). Suva, Fiji: University of the South Pacific Institute of Education.

Ahern, T. C., & Van Cleave, N. K. (2002). Utilizing open source and object oriented paradigms in instructional technology. *World Multiconference on Systemics, Cybernetics and Informatics.* Retrieved January 11, 2005, from http://www.ux1.eiu.edu/~cfnkv/Papers/SCI2002_OpenSource.pdf

Evers, V., Kukulska-Hulme, A., & Jones, A. (1999). Cross-cultural understanding of interface design: A cross-cultural analysis of icon recognition. In E. del Galdo & G. Prahbu (Eds.), *Proceedings of the International Workshop on Internationalisation of Products and Systems.* Rochester, NY. Retrieved August 9, 2004, from http://www.swi.psy.uva.nl/usr/evers/IWIPSFinal.pdf

Harris, S. (1992). "Going about it the right way" — Decolonising aboriginal school curriculum processes. In B. Teasdale & J. Teasdale (Eds.), *Voices in a seashell: Education, culture and identity* (pp. 37-53). Suva, Fiji: The University of the South Pacific Institute of Pacific Studies.

Hedberg, J. G., & Brown, I. (2002). Understanding cross-cultural meaning through visual media. *Education Media International, 39*(1), 23-30.

Kalolo, K. (2002). A Tokelau perspective. In F. Pene, A. Taufe'ulungaki, & C. Benson (Eds.), *Tree of opportunity: Rethinking Pacific education* (pp. 104–06). Suva, Fiji: University of the South Pacific Institute of Education.

Kisum, S. (2003). Selected country papers: Fiji (2). *Multimedia and e-learning: A new direction for productivity promotion and enhancement* (report of the APO Seminar on Multimedia for Productivity Promotion and Enhancement (with special focus on e-learning)). The Asian Productivity Organization, Japan. Retrieved January 12, 2005 from, http://www.apo-tokyo.org/00e-books/07.eLearning.htm

Landbeck, R., & Mugler, F. (1994). *Approaches to study and conceptions of learning of students at the USP.* Suva, Fiji: University of South Pacific Centre for the Enhancement of Learning and Teaching (CELT).

Lima, R. (2003). Educational ideas of the Noole (Solomon Islands). In K. H. Thaman (Ed.), *Educational ideas from Oceania: Selected readings* (pp. 120-125). Suva, Fiji: The University of the South Pacific Institute of Education, in association with the UNESCO Chair of Teacher Education and Culture.

Lockwood, F., Roberts, D. W., & Williams, I. (1998). Improving teaching at a distance within the University of the South Pacific. *International Journal of Educational Development, 8*, 265-270.

Matthewson, C. (1994). Whose development, whose needs? Distance education practice and politics in the South Pacific. *Journal of Distance Education, 9*(2), 35-47. Retrieved January 11, 2005, from http://cade.athabascau.ca/vol9.2/09_matthewson.html

McLoughlin, C. (1999). Culturally inclusive learning on the Web. *Proceedings of the teaching learning forum 1999,* The University of Western Australia (pp. 272-277). Retrieved January 12, 2005, from http://lsn.curtin.edu.au/tlf/tlf1999/mcloughlin.html

Mel, M. A. (2001). Interfacing global and indigenous knowledge in evaluation and assessment in information technologies. *Educational innovation for development: Interfacing global and indigenous knowledge*. Report of the Sixth UNESCO–ACEID International Conference on Education (pp. 61-66). Retrieved June 19, 2004, from www.unescobkk.org/ips/ebooks/%20documents/aceidconf6/themetwo.pdf

Pagram, J., Fetherston, T., & Rabbitt, E. (2000). Learning together: Using technology and authentic learning experiences to enhance the learning of distance education students. *Proceedings of the Australian Indigenous Education Conference,* Australia. Retrieved August 13, 2003, from http://www.kk.ecu.edu.au/sub/schoola/research/confs/aiec/papers/jpagram02.htm

Parrish, P. E. (2004). The trouble with learning objects. *Educational Technology Research and Development, 52*(1), 49-67.

Petaia, R. (1981). Kidnapped. *The Mana annual of creative writing, 1*(3). Suva, Fiji: Mana Publications.

Robbins, C. (2004). *Educational multimedia for the South Pacific*. Suva, Fiji: JICA ICT Capacity Building at USP Project. Retrieved January 11, 2005, from http://www.grographics.com/fiji/AccessUsability-ICTResearch/research-report/Education-Multimedia-for-the-South-Pacific_Robbins.pdf

Sami, G. (2004). Decolonizing our mind: Rationale, challenges and coping strategies. *Panel at the Fiji Institute for Educational Research.* The University of the South Pacific, Suva, Fiji, January 7, 2004.

Taafaki, I. (2001). *Effecting change and empowerment: Library/information studies training through distance education in the Marshall Islands* (Internal Document). Marshall Islands: University of the South Pacific Distance and Flexible Learning Centre.

Taufe'ulungaki, 'A. M. (2003). Vernacular languages and classroom interactions in the Pacific. In K. H. Thaman (Ed.), *Educational ideas from Oceania: Selected readings* (pp. 13-35). Suva, Fiji: The University of the South Pacific Institute of Education, in association with the UNESCO Chair of Teacher Education and Culture.

Thaman, K. H. (1992). Looking towards the source: A consideration of (cultural) context in teacher education. *Directions 27, 14*(2), 3-13.

Thaman, K. H. (1997). Considerations of culture in distance education in the Pacific Islands. In L. Rowan, L. Bartlett & T. Evans (Eds.), *Shifting borders: Globalisation, localisation and open and distance education* (pp. 23-43). Geelong: Deakin University Press.

Thaman, K. H. (1999, November). Of teachers and lecturers: Understanding students' cultures. *Centrepoint* (4) (pp. 1-2). Centre for the Enhancement of Learning and Teaching), The University of the South Pacific, Fiji Islands.

Thaman, K. H. (2000a, October 25-27). Open and flexible learning for whom? Rethinking distance education (Keynote address). *14th Annual Conference of the Asian Association of Open Universities (AAOU),* Manila, Philippines. Retrieved January 14, 2004, from http://www.friends-partners.org/utsumi/gu-l/early-2001/1-31-a.html

Thaman, K. H. (2000b). Decolonising Pacific studies: Indigenous knowledge and wisdom in higher education (Plenary address). *Pacific Studies 2000,* University of Hawaii at Manoa, Honolulu.

Thaman, K. H. (2003). A conceptual framework for analysing Pacific educational ideas: The case of Tonga. In K. H. Thaman (Ed.), *Educational ideas from Oceania: Selected readings* (pp. 73-8). Suva, Fiji: The University of the South Pacific Institute of Education, in association with the UNESCO Chair of Teacher Education and Culture.

UNESCO. (1992). Recommendations of the conference "Education for Cultural Development" Rarotonga. In B. Teasdale & J. Teasdale (Eds.), *Voices in a seashell: Education, culture and identity* (p. 7). Suva, Fiji: The University of the South Pacific Institute of Pacific Studies.

USP DFL Unit. (2004). *Distance and flexible learning at USP: About us.* Retrieved January 4, 2005, from http://www.usp.ac.fj/dfl/aboutus.htm

Va'a, R. (1997). Cultural accountability in the USP science courses at a distance. In L. Rowan, L. Bartlett, & T. Evans (Eds.), *Shifting borders: Globalisation, localisation and open and distance education* (pp. 83-97). Geelong: Deakin University Press.

Va'a. R. (2000). *Appropriate telecommunication and learning technology for distance education in the South Pacific* (Report of a project funded by NZODA). Suva, Fiji: Pacific Islands Regional Association for Distance Education, the University of the South Pacific.

Veramu, J. (2004, January 7). *Decolonizing our mind: Rationale, challenges and coping strategies.* Panel at the Fiji Institute for Educational Research, The University of the South Pacific, Suva, Fiji.

Wah, R. (1997). Distance education in the South Pacific: Issues and contradictions. In L. Rowan, L. Bartlett, & T. Evans (Eds.), *Shifting borders: Globalisation, localisation and open and distance education* (pp. 69-82). Geelong: Deakin University Press.

Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), *The instructional use of learning objects* [Online version]. Retrieved January 11, 2005, from http://reusability.org/read/chapters/wiley.doc

Yorston, F. (2002). *Learning to teach in Samoa. Master's of teaching internship case from four months in Samoa in 1999.* The University of Sydney, Australia. Retrieved January 15, 2004, from http://alex.edfac.usyd.edu.au/acrosscurric/Paradise/POP/centre_intern.html#An%20example%20of%20how%20effective

This work was previously published in Information Technology and Indigenous People, edited by L. E. Dyson, pp. 65-79, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 5.3
Using an Interactive Feedback Tool to Enhance Pronunciation in Language Learning

Felicia Zhang
University of Canberra, Australia

ABSTRACT

This chapter focuses on the effect of a learning environment in which biological, physical and technological ways of perceiving Mandarin Chinese sounds have been used. One of the most important tools of this environment is the use of a speech analysis tool for audio and visual feedback. This is done by way of incorporating a visual representation of student's production that can be easily compared to the speech of a native speaker. It is the contention of this chapter that such an interactive feedback tool in conjunction with other feedback mechanisms can provide opportunities for increasing the effectiveness of feedback in language learning.

INTRODUCTION

This chapter reports on an experiment of restructuring the learning environment using a variety of computer-enhanced language tools with the explicit aim of training students to perceive Mandarin (hereafter referred to as "MC") sounds. It focuses on the effect of creating a learning environment in which biological, physical, and technological ways of perceiving MC sound have been taught to students. It is hoped that access to these different approaches to perception will help students to know how to better perceive MC sounds outside the classroom context. One of the most important parts of this environment is the use of a speech analysis tool for offering audio and visual feedback by way of incorporating a visual representation of student's production that can be easily compared to the speech of a native speaker. It is the contention of this chapter that an interactive feedback tool such as this can provide opportunities for increasing the effectiveness of feedback in language learning. It is hoped that through the exploration of the results of this research, clearer directions on how this technology can be generalized to other learning contexts with other languages can emerge.

CRITIQUE OF VARIOUS WAYS OF TEACHING PRONUNCIATION

Practitioners of both "traditional" and "modern" approaches of language teaching have generally acknowledged good pronunciation as a very important objective in learning a second language (L2). As perception is intricately connected to speech production, training to perceive sounds necessarily becomes an important part of language acquisition and good pronunciation acquisition. However, in the history of foreign language instruction, pronunciation has not always been regarded in this light.

The grammar translation method, which focuses almost entirely upon written texts, has always considered pronunciation nearly irrelevant. The cognitive code approach also de-emphasized pronunciation in favor of grammar and vocabulary, because it was thought in the late 1960s and early 1970s that native-like pronunciation could not be taught anyway (Scovel, 1969).

Subsequent approaches, however, put more emphasis on oral communication. For example, the direct method has claimed that pronunciation is very important and presents it via teacher modeling. This methodology assumes that sounds practiced in chorus or even individually will automatically be transformed into "correct" production by the students. Similarly, the immersion method assumed that students would acquire good pronunciation through exposure. In the audiolingual approach, pronunciation is also very important. In this approach, the teacher models and the students repeat, usually with the help of minimal pair drills. However, by making students "improve" their pronunciation through a set of minimal pair drills suggests that every learner will make a particular error through a particular trajectory. For example, if an English as a foreign language (EFL) learner makes an error with the word "beach," it will inevitably be that he or she will say it as "bitch." This predetermination of what kind of errors students will make when learning a L2, not only denies a student's individuality, it also excludes many other possible causes that may lead a learner to make that particular error.

Even in the teaching approaches that focus on oral communication cited above, relatively scanty attention has been paid to the complex nature of phonation and auditory perception in a L2. In fact, teaching people to perceive in a L2 is considered so difficult that most teaching methodologies have based their approaches for teaching pronunciation on the teaching of elements that are relatively easy to define (e.g., vowel and consonant sounds). Elements that are relatively unstable and hard to define, such as intonation patterns, are usually left out of the teaching process. The teaching of intonation and rhythm has hardly been explored. The logic behind this is easily understood: one must first put together the elements of language and then, later, somehow add the intonation. These methods of teaching pronunciation have been widely used in language teaching. However, they have not yielded particularly useful results, for instance, in the field of teaching English as a second/foreign language. This led Jenkins (Jenkins, 2000) to argue that in the case of the English language, as many nations in the world use a variety of English as their own native or official language, rather than measuring native-like pronunciation or intelligibility against any particular form of the English language from say, the United Kingdom or the United States of America, it might be worthwhile to set up an international core for phonological intelligibility for the English language. It suffices to say that within this core for phonological intelligibility for English, prosodic features such as intonation were the least important according to Jenkins' reasoning. Yet this might, in fact, be approaching the problem from the wrong direction entirely.

Moreover, one cannot possibly dismiss the relationship between good pronunciation and social power. If one wants to be accepted and respected in the target language culture, the first

testament of one's worth is one's pronunciation and fluency in that particular target language. Thus, mastery of intonation patterns of that L2 is actually an integral and crucial part of language proficiency.

Finally, research by Hinofotis and Bailey (Hinofotis & Bailey, 1980) has demonstrated that "there is a threshold level of pronunciation in English such that if a given non-native speaker's pronunciation falls below this level, he or she will not be able to communicate orally no matter how good his or her control of English grammar and vocabulary might be." It is then reasonable to assume that there might also be a similar threshold level of pronunciation in MC for non-native speakers of MC. Tones in MC, as an indispensable part of intonation, perform the function of differentiating word meaning. The importance and endurance of tones in MC is such that native speakers will still find your speech intelligible, even if the vowels and consonants are unintelligible (Chao, 1972). This is why recognition bestowed upon a non-native speaker's mastery of MC is almost entirely based on native speakers' perceptions of their tones. In other words, intelligibility of non-native MC is based very much on the speaker's correct tonal production. For this reason, mastery of tone proves to be one of the most worthwhile tasks in learning spoken MC. This is true not only of a tonal language such as MC; it is also true of a nontonal language such as French (James, 1976).

In short, it is maintained that instead of focusing on the easily definable and discrete elements that make up speech, perhaps a worthwhile experiment would be to start with the intonation and the melody of a language and the process of training L2 students to perceive sounds in a L2. The study described in this chapter reports on an experiment based on this orientation of pronunciation training.

OBJECTIVES OF FEEDBACK

Getting good-quality feedback is an important aspect of language learning (and learning in general). For many educationalists, feedback is important, because it is essential for learning and can play a significant role in students' development by providing knowledge required for improvement (Hinett, 1998; Hyland, 2000). The objectives of providing feedback are as follows:

1. To enable students to understand feedback and to make sense of it;
2. To establish a common understanding of how this feedback may be implemented or acted upon by students between students and teachers.

Toohey (2000, p.154) gave a model of a learning process involving feedback: initially the student encounters or is introduced to an idea, this is followed by the student becoming aware of the idea, the student then tries the idea out, receives feedback and then reflects and adjusts the implementation of the idea. Feedback as described in this model takes place in a language teaching classroom on a daily basis. In a communicative classroom, students are frequently called upon or volunteer to try out a new sentence, conversation, or structures. The student then receives feedback from the teacher or his or her fellow classmates almost instantly in many ways. In the classroom, many channels are used for communicating feedback, through error correction and other channels, such as body language, nonverbal behavior, facial expressions, gestures, tone of voice, and so on. Such feedback is usually instantaneous, involuntary (from the feedback provider), episodic, and disappears very quickly from the memory of everyone involved. So in a traditional classroom, while we receive a huge amount of feedback on our production, the feedback received seldom becomes guidance or long-term learning in a real,

face-to-face communicative interaction outside the classroom.

First, can we expect the classroom situation to provide feedback that is able to achieve the objectives listed above? Clearly, the episodic nature of the feedback offered in the classroom can have minimal effect on student learning. Second, understanding what the feedback contains and how to act on it are not as easy as they seem. The feedback offered in a L2 language classroom does not only contain information on the correct way of pronouncing or writing something. In some cases, error correction offered can be as detrimental as not offering any feedback at all. An excellent example is in the teaching of MC tones. In a character-based language such as MC, each character has a lexical tone that is stable when it is isolated from other words. However, in a sentence situation, a lexical tone of a character is influenced by other characters and their tones before and after it. In other words, a character might lose its stable lexical tone in a larger stretch of discourse. In the MC language, changes in tones across sentences and longer stretches of discourse are very hard to predict and describe. Yet, in every classroom in the world where students are learning MC as a foreign language, teachers are constantly pulling students up on their tones by demonstrating the right tones for individual characters and telling them the tone for a particular character is a fourth tone not a first tone and so on. The problem is that such corrective effort is usually ineffective, as the immediate context (influence of words left and right to the character in question) of the correction for that particular character is ignored, and the effect is usually short-lived and transitory.

THE CONTRIBUTION OF COMPUTER TECHNOLOGY IN FEEDBACK PROVISION

The advent of computer technology in language learning has added a very interesting dimension to the role of providing feedback. Computer technology can provide an environment in which certain memory traces that work for a particular learner can stay longer in that students' consciousness or sometimes unconsciousness. In a real-life situation, such memory traces can be called upon to help facilitate the communication involved. The advantages of feedback offered by a computer are that the feedback is constant; it can be repeated over and over; and it allows students to control their own learning.

However, the design of CALL tools for providing feedback still depends on the theoretical framework that conceptualizes the tools. For instance, in a Speech-Recognition-Based pronunciation training experiment conducted by Tomokiyo et al. (2000), the feedback offered still follows the model of offering explanation using diagrams of articulatory position and minimal pair practice. Many CALL tools still tend to focus on the production side of pronunciation rather than on exploring how students perceive sounds. Thus, the criticisms that are put forward against the various approaches of teaching pronunciation in language teaching are equally pertinent here as well. The present study proposes that instead of conducting research based on the more stable aspects of pronunciation (i.e., vowels and consonants) that are constrained by particular theories of linguistics, it might be more productive for students if we conduct research based on sound theories of learning. The speech tool we developed, and which will be discussed in more detail shortly, is based on verbo-tonal system of phonetic correction, developed by the late Professor Guberina in connection with his work with people whose hearing was impaired (Renard, 1975). This is a system that brings the human brain and the human body as a whole to the forefront of the study of auditory perception. The computer software is used in a language learning environment that has been designed to use all the sensory organs of the body to facilitate auditory, visual, and other perceptions and to contribute to the brain's realization of every perception.

Most of the feedback that is currently offered in computerized multimedia environments is focused on the product of student's performance rather than upon the process. For instance, we are familiar with many language-teaching exercise makers that allow you to offer feedback such as "try again" or "well done." However, such feedback is pointless, and students are unlikely to benefit from it if they do not change their actions during the process of production. In the case of getting better pronunciation in any language, this amounts to getting students to try to say a sentence differently. Most feedback mechanisms, while offering feedback in terms of judgment, do not offer any feedback that contributes to the process of production. The speech tool we developed, by contrast, is designed to offer feedback that is nonjudgmental and allows students to explore and reflect during the process of learning, not just at the end of the learning process. Reflection occurs when students can observe visually the differences between their productions and the native speaker model. When this is combined with the biological and physical memory traces built up in the classroom context, students can act upon their reflections and change the processes of production. It is the contention that this way, students will be able to turn feedback to students into learning in the long run.

Providing Feedback Using Sptool

A few attempts have been made to teach melody and intonation in a new and original way by displaying the melodic contours on a screen. The students hear a model sentence and, at the same time, they watch the melodic pattern form on the screen where it is "frozen." They then attempt to match the model melody by speaking. As their melodic pattern is also "frozen," they can compare their productions with the model. Students can then employ auditory and visual stimuli for assistance in the comparison process. Such methods, overall, have had good results (James, 1976). In the past, this kind of feedback machine was relatively expensive and, as only one person could be trained on any single machine at any one time, not really feasible for widespread educational use. However, computer technology has made it possible that such a feedback mechanism can be provided at the click of a button. The Sptool (Zhang & Newman, 2003) used in this research is such a feedback mechanism.

The Sptool program is produced with a windows component called dotnetfx.exe; it is 20 megabytes in size. It is most stable for Windows 2000 and XP; a bit unstable for Windows 98. When you open the program, you can see the open icon, which allows you to import any other prerecorded audio file saved in the .chwav, .mchwav, or .wav formats. If you have a prerecorded sound file, saved in Windows .wav format, you should be able to open it and run through the program. If you record a male voice in Cooledit (Syntrillium, 2002) or any other recording software, save it under the Windows .wav format to .mchwav. If you record a female voice, save it to .chwav, and then Sptool should play and measure your recorded sentence. At the moment, the program is not able to work with large files over 10 megabytes due to the limitation imposed by the Microsoft component.

Other Speech Analysis Tools

There are other speech analysis tools on the market. The commercial product Winpitch (Martin, 2003) is such a speech analysis tool. Praat (Boersma & Weenink, 2003) is free and also available. However, these software programs are far too complex for beginning language students to use. As the students involved in this research are zero-level beginners of MC with varying computer literacy levels, the complexity of the programs described above made the creation of a more user-friendly speech analysis tool integrated into the teaching material a necessity.

The Tell-Me-More series (Auralog, 2000) of language-learning programs also have built within them a speech comparison tool. However, the speech comparison chosen by the creator of Tell-Me-More is using phoneme matching. In other words, while the sentences contained in the language program can be verified by the speech comparison tool attached to the program, it cannot verify any other sentences outside the program. This means the usefulness of their speech tool is really limited.

THEORETICAL FRAMEWORK UNDERPINNING THE CURRENT RESEARCH

The use of Sptool is embedded in a larger learning theory based on the theory of the verbo-tonal system of phonetic correction. This theory is mainly concerned with the way students perceive sounds of a L2. The starting point of this theory is the complex nature of phonation and auditory perception of a language. From the verbo-tonal point of view, auditory perception develops synchronously and synergistically with the development of the motor, proprioceptive, and visual abilities (Guberina, 1985). One of the senses in audition is through the ears. A person with normal hearing in his mother tongue will behave, in a foreign language, as though he or she were hard of hearing. Each language sound carries all frequencies from about 50 Hz to about 16,000 Hz (albeit at various intensities). Theoretically, at any rate, each sound can be heard in many different ways. The ear seems to have a "choice" as to what to hear, in practice, depending on the way in which the ear has been trained. L2 students tend to make "choices" in the target language based on what they are familiar with in their mother tongue. Each sound has a particular "optimal" frequency (i.e., the frequency band, or combination of frequency bands, at which a native speaker best recognizes and perceives the sound in question). This is what Troubetzkoy (1969) referred to as the mother tongue "sieve."

Students who experience difficulty with a particular foreign language sound are considered as not having recognized its optimal. Hence, they are unable to reproduce the sound correctly. One of the ways in which students can be made to perceive the optimal of each sound is to remove (e.g., through electronic filtering) any interfering frequencies that might prevent it from being perceived. In this way, it is possible, in theory, to bypass the mother tongue "sieve" (Troubetzkoy, 1969). Once this has been achieved, students will be able to perceive, for the first time, the specific quality of the troublesome sound. However, exposing the students to the native speaker optimal may still be insufficient. A set of "corrective" optimals then needs to be determined. These will be such as to direct a student's audition away from its natural tendency to structure as it has always done.

Verbo-tonalism postulates that the articulation of sounds poses relatively little difficulty once the specific quality of the sound has been heard. Consequently, the determination of corrective optimals for any one student will be established on the basis of his or her pronunciation. It is through exposure to corrective optimals, followed by intensive articulatory practice, that students will carry with them valid acoustic models constituting the normal range for the phonemes of language. The intense exposures to the sentences in this course via the Sptool plus the intensive articulatory practice carried out in the two-hour lecture provide students with such valid acoustic models of the phonemes of the L2.

Audition is a form of total behavior that occurs on the level of the body as a whole. In the present course, nine steps in the lecture sequence have been designed to integrate phonation and expressiveness that occur in the space between the lungs and the nasal cavity, with the breathing, moving, feeling patterns of a person in entirety so that a multitude of memory traces will be retained in different parts of the body.

Given the complexity of the various processes involved in perception and phonation, an intellectualization of these processes is unlikely to be successful. Learning processes must therefore operate at the "unconscious" level. Rather, it is essential that proprioceptive powers be called into play in the development of good pronunciation so that students might become conscious and perceptive of the rhythms and stresses of the target language. The fact that translation into English, romanization in Pinyin, etc. are not emphasized or used at all in this course suggests that the course is especially designed to allow new language to be processed "unconsciously"—or perhaps *intuitively* would be a better word—first and foremost. In other words, we are not really talking about unconscious learning but the more intensive utilization of the language centers through the exploitation of different parts of the nervous system, such as the parts that are concerned with proprioception and bodily sensation.

The elements described in the following lecture sequence and the audiovisual materials contained in the teaching materials represent the pedagogic measures that integrate the senses of the body with movement with the process of ear training through working on a system of errors rather than isolated elements of the language. It is proposed that starting an audition process from intonation would result in the proper training of several systems at once in MC. These pedagogic measures also are designed to instill in students certain memory traces by physically "marking" on their brains so that these memory traces can be reactivated once feedback either from Sptool or from any other sources has been received.

These memory traces are essential in enabling students to act upon the feedback received. This is the second important objective of any feedback system: to create a set of memories that are not merely cognitive records but feelings of relaxation and muscular tension that are distributed through the students' experience of his or her bodily sense. What follows is a brief description of a teaching method that, in helping to create such somatic traces, provides an environment suitable to the inclusion of the Sptool and assists students in the exploitation of that resource.

A NEW METHOD OF TEACHING MANDARIN CHINESE PRONUNCIATION TO BEGINNERS

It is 5:30 pm on a Tuesday afternoon, in a large room capable of holding up to 50 students; the lecture chairs with attached arms have been pushed to the perimeter of the room. The students are randomly slouched on their chairs relaxing after a tired day of either work or lectures.

The teacher walks into the room carrying the necessary computer gear, CD-ROMs and so on. She greets the class cheerfully with "ni3men hao3" (hello, everyone) and puts the CD-ROM in the computer. "Now, leave your seat and lie comfortably on the floor and listen." Then the following audio file is played:

"Imagine that you are lying on your back on the grass on a warm summer day and that you are watching the clear blue sky without a single cloud in it (pause). You are lying very comfortably, you are very relaxed and happy (pause). You are simply enjoying the experience of watching the clear, beautiful blue sky (pause). As you are lying there, completely relaxed, enjoying yourself (pause), far off on the horizon you notice a tiny white cloud (pause). You are fascinated by the simple beauty of the small white cloud against the clear blue sky (pause). The little white cloud starts to move slowly toward you (pause). You are laying there, completely relaxed, very much at peace with yourself, watching the little white cloud drift slowly toward you (pause). The little white cloud drifts slowly toward you (pause). You are enjoying the beauty of the clear blue sky and the little white cloud (pause). Finally the little white cloud comes to a stop overhead (pause). Completely relaxed,

you are enjoying this beautiful scene (pause). You are very relaxed, very much at peace with yourself, and simply enjoying the beauty of the little white cloud in the blue sky (pause). Now become the little white cloud. Project yourself into it (pause). You are the little white cloud, completely diffused, puffy, relaxed, very much at peace with yourself (pause). Now you are completely relaxed, your mind is completely calm (pause), you are pleasantly relaxed, ready to proceed with the lesson (pause)." ***(Step 1)***

"Now, get up and stand in a circle." The teacher joins the circle.

The teacher says "I will hum to the rhythm of the sentence and please hum with me while walking slowly in a circle." This is done five times. ***(Step 2)***

"Now, I will clap to the rhythm of the sentence and then you can clap after me." ***(Step 3)*** *Again, this is done five times.*

"Notice the high sounds and the low sounds in the sentence? With your palm up, push your hands above your head as high as possible for the high sounds. For the low sounds, stamp your feet done as hard as possible. Now let's hum the sentences again." ***(Step 4)*** *This is again done five times.*

"Continuing with the movements, now mouth the sentences while I say them out loud." ***(Step 5)*** *Of course, at every lesson, at this stage, one or two people always end up repeating the sentences rather than mouthing them. This is again done five times.*

"Now repeat after me, and then add words to the intonation." This again is done five times. ***(Step 6)***

Now the teacher instructs each individual to repeat the sentence by themselves; checking that each student is reproducing the sentence correctly. ***(Step 7)***

"Now what is the meaning of the sentence?" Students enthusiastically volunteer the meaning in English, and the meaning of the sentence is usually established in seconds. ***(Step 8)***

In each two-hour lecture sequence, every sentence is presented and practiced using the above procedure. At the end of each lecture, the whole class engages in a pair or group work conversation activity using the materials covered in the lesson. ***(Step 9)***

At the end of the lesson, students are instructed to sit and write the meaning or whatever notes they want to make themselves.

In the lecture sequence described, several rather unconventional elements make an appearance. For instance, relaxation exercise, humming, mouthing to the words, body gestures, and mouthing the words and then repetition, are all present in the learning sequence. How are these related to each other? How do these elements relate to the Sptool under discussion?

Focusing on the Rhythm and Intonation of the Language

The activities described in the above lecture sequence all have to do with focusing on the rhythm and intonation of the language. Intonation is a universal feature of all languages. Melody (which includes tones and intonation) holds the units together and arranges them with respect to one another. It is a very special kind of glue. Attempting to arrange the sounds with the wrong "glue" is like building structures of the wrong kind.

The smallest unit of the language being presented is a sentence rather than individual words or compounds. This is because in MC, the acoustic characteristics of the words change when they are in a sentential environment. For instance, when a word is read in isolation, the frequency of the word is different from when the word is part of a sentence. So concentrating one's effort in mastering the tones of individual words or compounds does not guarantee success in producing the sentences containing those words. This is true of MC as well as other nontonal languages.

Step 1: In this step, the imagery of the "little white cloud" is used to relax the students and the teacher. This constitutes the *relaxation* phase. This sets up a relaxed atmosphere for learning for the rest of the lesson. Stevick (1986) has stressed the usefulness of working with imagery in language teaching. In education, visualization can facilitate the interiorization of knowledge by creating a more receptive state of awareness, permitting the affective and creative functions of a more holistic nature to participate in and strengthen the learning experience (Murdock, 1987). According to Neville (Neville, 1989), "the fragmented, dispersed, chaotic energies of our organism are aligned, harmonized and made purposive by the imagined experience, just as they would be by a 'real one', possibly leading to important changes in our 'self-image, attitude and behavior' (p. 95).

Step 2: This step involves humming along to the rhythm of the sentences without the vowels and consonants (five times). This is used to highlight the intonation and tones of MC.

Step 3: Clapping to the rhythm allows students to experience the rhythm of the sentence and observe different groupings of the words in a sentence. This also allows the students to observe how stress, realized by length and loudness in MC, is tied to meaning. This also allows them to observe the key words in a sentence and realize that not all words are of equal value and that in making oneself understood, one only needs to get the key words right. This training is essential in training them the strategy of prediction and advanced planning in listening comprehension.

Step 4: In this step, walking about with feet coming down on every syllable is practiced in order to get the body used to producing a tense downward tone that is also loud. Raising or stretching upwards as though attempting to touch the ceiling to experience the tenseness of the body in producing the first high level tone is also done. Instruct the students to adopt a forward lumping of the shoulders for the second and third tones in MC that need a relaxed posture.

Research shows that Chinese speakers have a much wider voice range when speaking MC than English speakers speaking English (Chen, 1974). As the first tone starts at a higher frequency than what most Australian speakers are used to, extra physical efforts need to be made to remind one that one must start high. To stretch one's muscular system to express these MC tones, one must not slouch in seats. By asking students to stand up straight and walk in a circle with various gestures, students are experiencing the coordination and synchronization of various muscles with the sounds uttered.

Steps 5 through 9 are steps that further highlight the melody of the sentences involved. Notice that throughout the learning sequence, translation and writing down the sentences are not needed until the last moment. By the time students come to write down the meaning, they will have already internalized the melody of the sentences.

The nine steps of the lecture sequence offer students a range of physical ways for remembering tones beyond the set contact hours every week. These measures set up a series of learning steps that can be used for self-access learning at home.

ROLE OF THE SPEECH TOOL AND THE COURSE DATA CD AND AUDIO CD

Course Data CD

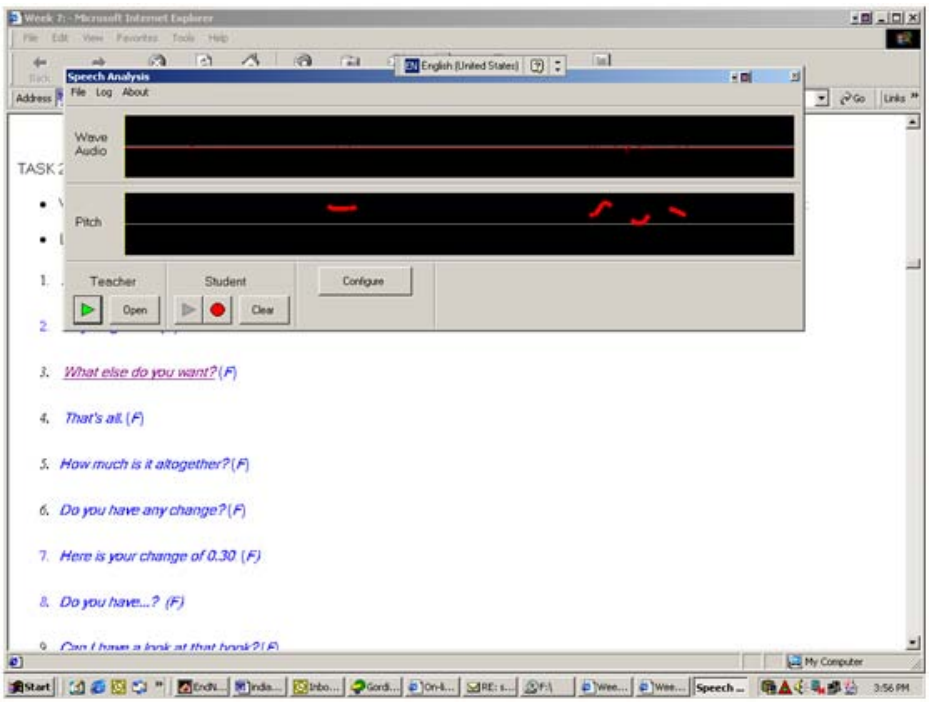
Each new vocabulary item, new sentence, or new phrase in the teaching materials is linked to a normal sound file. Only the sentences are linked to both a normal sound file and a filtered sound file.

All the sound files in the materials can be passed through the Sptool (Figure 1). Once passed through the Sptool, the learner can listen to the teacher's model pronunciation by clicking on the "teacher" icon. With one click, the student can hear the model sentence, and the pitch curves of the model sentence are displayed. If the learner wants to hear a smaller chunk of the sentence, then he or she can select that bit of the curve by dragging the cursor over the portion they want to hear. After listening to the sentence numerous times, the learner can decide whether he or she wants to record his or her own production.

Before clicking on the record button, however, it is necessary for the student to tell the program whether he or she is female or male. This is necessary because females generally have higher fundamental frequencies than males. The difference is sometimes as much as over 100 Hz. If the program is configured to measure a female voice, the pitch curve of a male learner will not be able to be displayed. However, once configured correctly, the pitch curve of the learner's recording will then be displayed properly.

The course data CD also contains teaching materials in html format; all the associated sound files, the Speech tool, and short video skits with which students can test their comprehensions of the new language can be learned by watching these video skits. An audio CD of the sound files is also provided with the course materials. In 2003,

Figure 1. Picture of the teaching material



a weekly compulsory class using computer-enhanced teaching materials was arranged. Other computer-enhanced learning materials such as Tell-Me-More (Auralog, 2000), VCDs, movies, and videos with similar content are also available in the computer center.

The Role of Sptool

Steps 2 to 7 in the lecture sequence are duplicated in different forms through the use of the Sptool. While the classroom sequence is more or less teacher driven and physical, the Sptool allows the lecture sequence to be experienced in a different way. It also allows other choices to be made:

- The beat, stress, word groupings, key words, and sentential intonation are all indicated in the speech curve and the sound file. In the sample sentence, zai4 nar3 you3 mai4 bi3 de0 (Where can one buy some pens?) shown in Figure 2 on the next page, "zai4 nar3" is the key word and the curve clearly shows that the three characters are in a group together and should not be separated.
- The height (related to the muscular tenseness of the body) of both first and fourth tones is indicated clearly with respect to other tones. The height of the first and fourth tones reminds the students of the need to stretch their voice range beyond their normal voice range. This information is very useful in enabling students to change their ways of producing the target sentences after observing the differences between the native speaker's production and theirs.
- Comparison of the pitch curves of individual words in the vocabulary section with the same words used in sentences is possible.
- Students can select any portion of the sentence for listening practice and repetition.
- The links between words are easily observable. For instance, in the sentence,
  Wo3 sheng1yu2 yi1 jiu3 wu3 ling2 nian2.
  I was born in the year 1950.

The production of "sheng1yu2" requires the body to be tense and to be kept tense in order to produce the next "yi1: one." Students can select the three syllables "sheng1yu2yi1" in order to explore how physically one has to keep one's body tense in order to produce this group of words using the physical gestures practiced in the classroom.

The use of Sptool encourages students to reflect on and explore the process of learning. Many of the explorations are usually impossible to be predetermined by a teacher, as most teachers, even the most able, do not have an extensive list of rules about how the different combination of words are produced physically in MC. Many of the things that can be done using the program may not be initiated by teachers but are being explored by the students through use. Furthermore, being able to experience each sentence repeatedly through the Sptool creates an environment in which students can totally immerse themselves consciously and unconsciously in the language.

THE STUDY

Sample

The progress of three groups of beginners of Mandarin Chinese has been followed. The first group (hereafter referred to as "Group 1") consists of two other groups of total beginners from 1995 and 1996 who were taught pinyin (the Chinese romanization) from the beginning of their MC study. The oral test data collected from this group represent the baseline data.

The second group (hereafter referred to as "Group 2") of beginners finished their two semesters of study in Mandarin in 2001. These participants were students enrolled in *Chinese 1a: Language and Culture* and *Chinese 1b* of the first-year Chinese course at the University of Canberra in 2001. By the end of the experiment, they had completed 130 contact hours of lectures and tutorials over two semesters.

The third group (hereafter referred to as "Group 3") consisted of students who studied MC in the first semester, 2003. Students in Group 3 were zero-level beginners when they started and were taught exclusively by the use of the Sptool. These participants in this study were 15 students enrolled in *Chinese 1a: Language and Culture* in 2003. There were three students from Japan, one student from Korea, and 10 Australian students. By the end of the experiment, they would have completed 65 contact hours of lectures and tutorials over one semester. They were all zero-level beginners of MC at the beginning of the course.

Data Collection Methods

A configuration of data methods was used to explore the experiences of Groups 2 and 3 students as they learned MC through this technology-rich learning environment. The configuration of methods is as follows:

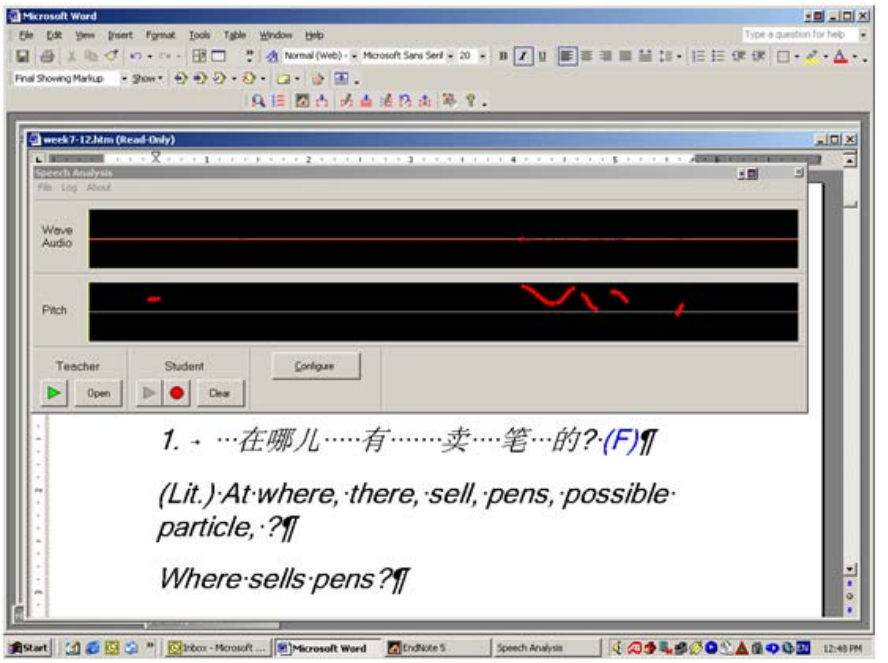
Qualitative:

- One-to-one oral tests (four of them for Group 2; one for Group 3)
- Self-scripted video segment done in small groups (four for Groups 1 and 2; one for Group 3)
- Written examination tests (four for Group 2; and one for Group 3)

Quantitative:

- Computer-technology-related questionnaires from Group 2 students
- One-to-one interviews with Group 2 students

Figure 2: Picture of the Sptool showing the sample sentence: Where does it sell pens?



RESULTS AND DISCUSSION

Qualitative Data

Group 2:

Data collected from the 2001 group of students were compared with two other groups of total beginners from 1995 and 1996. The fundamental difference between the groups was that the 2001 group of students was not taught pinyin from the very beginning. Analysis of students' oral performances revealed that the rate of acquisition of MC initials (consonants) and finals (vowels) was faster for these students. For instance, in the data from students in Group 1 (the pinyin group), a large number of errors with palatals [x] and [j] and [q] were present. While the Group 2 students made some errors with [x] and [q], errors with [j], after only six weeks of instruction, did not occur. Furthermore, by the time the second oral (after 65 hours of instruction) was conducted, no errors were made with respect to these initials.

Group 3:

Students in Group 3 started learning Mandarin only since the end of February 2003. This group is the only group that benefited with a verbo-tonal-theory-driven teaching methodology combined with a fully developed software package that included course data and audio CDs and the speech tool. By the end of May 2003, this group of students had completed their first written and oral tests.

Oral test:

The first oral performances of eight students in Group 3 had been analyzed. Out of 1827 words (MC characters) uttered, 77 errors were made with consonants and vowels. In other words, only 4.2% of errors were made. Out of the 77 errors, students from a non-Australian background, i.e., one Korean, one Thai, and one Cantonese speaker of Chinese made the majority of the errors. In terms of consonants, only one Australian student made two errors with "zhi3." He pronounced it as "zi3." Again, errors with palatals [x] and [j] and [q] were not present. After only 30 hours of face-to-face instruction, all the Australian students had gained complete control of the initials of MC. Similar to previous groups of students, pinyin and non-pinyin groups, the problematic errors were with the vowels and diphthongs rather than with consonants.

The faster rate of Mandarin sound system acquisition can be attributed to the removal of romanization and the availability of sound files on CDs and the speech tool. The combined effect of both tools appeared to have helped to reduce the transfer effect from the students' mother tongue—English. A close examination of the audio recordings of Group 3 students' oral also suggests that Group 3 students were more fluent. Though they still had tone problems, the rhythm of their speech was much more natural when compared with a native speaker's rhythm.

Furthermore, anecdotal reports from students suggested that students were confident with their listening comprehension ability, citing that the use of humming, clapping, and so on in class combined with the use of Sptool, allowed them to hear and remember more distinctly the rhythm and stress of the language. This meant that they were able to hear familiar key words that enabled them to predicate what was coming up more accurately. Evidence of this came in the form of the ranking of students in different parts of their written tests. For this group, the top three ranks were occupied by Australian zero-level beginners overall. For the entire group, for the listening comprehension section, even the weakest student (the student who scored the lowest for the entire written test) scored well for this section. This was very different from the results of previous years. As one student remarked, this course was more like an immersion language program in which

students were expected to use the whole body in the process of learning.

Quantitative Data

Group 2: Results of the Computer Technology Questionnaires:

The reaction to the computer-based materials was extremely positive with this group of students. Many, 85%, of the students regularly used the audio and data CDs for at least one to three hours per week in an evaluation survey conducted at the end of the first semester in 2001. At the end of the first semester, students requested that in the written text on the data CD, each line of each dialogue be linked to its corresponding sound file, thus making the practice of the target language, line by line, easier.

A similar technology-oriented questionnaire was also administered at the end of the second semester. All students used the course data and audio CDs regularly to prepare and review the materials covered. Two out of six students requested each item of the vocabulary each week to be linked to its own audio file. Two students requested more regular use of VCD, and some requested more regular use of the Tell-Me-More CD in class.

Group 3:

No quantitative data had been collected from this group of students at the time of writing this chapter. Both questionnaire and interview data have been collected in late June of 2003 but are yet to be analyzed. A brief glimpse of the questionnaires completed by Group 3 students suggests that, on average, these students spent around 10 hours per person outside class contact time on their Chinese learning. This kind of devotion to the learning of MC has not been experienced by the researcher in her entire teaching career.

The availability of such an array of computer-enhanced learning materials encouraged students to engage in more autonomous learning behavior. As pronunciation could only be obtained either through looking in the dictionary or listening to the accompanying course CDs, all students spent time on a regular basis to listen to the CDs to prepare for the week's materials. This autonomous learning pattern actually forced students to open their ears to the target language in and outside of class, thus enormously increasing exposure to the target language. Autonomous behaviors happen every time each student listens to a string of sounds in Mandarin, as each student has to perceive the sounds according to his or her individual perception and translate them in a way that is recognizable to each student, individually.

FUTURE DIRECTIONS

It is important to note that Sptool can be used with any language, whether mother tongue, L2, or languages in danger of becoming extinct. It can be used with languages or any sound wave. Therefore, it can be used to enhance the teaching of the prosodic aspect of any language. For instance, in the teaching of English, the Sptool described in this chapter makes it possible for different models of native English sentences to be made available to teachers and students.

The Sptool used in the learning process is by no means perfect and, therefore, can be further improved in several aspects:

1. At the moment, it can only work with pre-produced sound files of a fairly small size. This means that a sound file has to be prerecorded before it can be run through this tool. It will be a huge improvement if as soon as someone speaks, the speech is recorded and automatically analyzed with the pitch wave displayed on the screen instantly. In other words, this tool should be able to process speech in real time.

2. Another icon called "filtered" can be built into the program to display the filtered version of any sentences.
3. This program should be able to talk to other programs, such as a video database, so that the sound file from any movie can also be measured and displayed instantly.
4. Band passes can be built into the program so that students can investigate the "optimal" frequency of a particular consonant or vowel.
5. A Web-version of the course should be created, and flash technology should be used so that once students point on a sound file, the speech tool opens automatically.

While Improvements 1 through 3 will not make the program too complicated to use, a large amount of research needs to go into investigating the possibility of Improvement 4. Improvement 5 is also possible, but the use of the World Wide Web itself already restricts its use. However, a Web version of the Sptool is being planned.

One of the most promising research directions at the moment with regards to this program is further improvement of the program and then testing of the tool with a larger group of students in different languages.

CONCLUSION

Limited findings described in this article demonstrated that a well-thought-out and properly implemented curriculum involving computer technology can make feedback to students more effective. This kind of environment is instrumental to produce students with better pronunciation in a L2 and can increase students' motivation for learning the language and culture of an L2. One significant consideration of creating this environment is the fact that the technology chosen is of a "*low-tech*" nature, utilizing mainly CD-ROM technology. While it has been acknowledged that adding the Net to this learning process may also be beneficial for students, at the beginning stage of language learning, the use of the Net only serves to increase the cognitive load on students.

Another significant characteristic of this environment is its *modularity*. I would like to refer to all the elements within the environment metaphorically as "machines." The "machines" used in the environment are easily accessible and user friendly and adaptable. The non-technology-driven elements such as one's body, voice, movement, and gesture are already available to every student. The technological elements, such as the sound, video, text files, and filtered sound files, are not hard to produce. The frequency of interaction and ease of access afforded by the Sptool and other sound files have been extremely motivating for students. Through the use of various machines in this environment, feedback, offered through physical, biological, and technological means, has acted in concert to motivate students and convert feedback in learning.

ACKNOWLEDGMENT

The investigation described in this chapter is sponsored by a University of Canberra research grant, 2002–2003 from the University of Canberra, Australia. I would like to thank Professor Michael Wagner for participating in the grant and offering me useful feedback and advice throughout the grant.

I would also like to thank Kate Wilson and two anonymous reviewers for helpful feedback on a previous version of this chapter. I am, however, entirely responsible for the good and bad herein.

REFERENCES

Auralog. (2000). Tell Me More (Asian). Auralog S.A.

Boersma, P. A., & Weenink, D. (2003). Praat. Institute of Phonetic Sciences, Institute of Phonetic Sciences, University of Amsterdam.

Chao, Y. R. (1972). *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press.

Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics, 2*(2), 159–171.

Guberina, P. (1985). The role of the body in learning foreign languages. *R. P. A., 73, 74, 75*, pp. 38–50.

Hinett, K. (1998). *The role of dialogue and self assessment in improving student learning.* British Educational Research Association Annual Conference, The Queen's University of Belfast.

Hinofotis, F., & Bailey, K. (1980). American undergraduates' reactions to the communication skills of foreign teaching assistants. In J. C. Fisher, M. A. Clarke, & J. Schacter (Eds.), *TESOL '80* (pp. 120–133), Washington, DC, Teachers of English to speakers of other languages.

Hyland, F. (2000). ESL writers and feedback: Giving more autonomy to students. *Language Teaching Research, 4*(1), 33–54.

James, E. F. (1976). The acquisition of prosodic features using a speech visualiser. *International Review of Applied Linguistics in Language Teaching, 14*, pp. 227–243.

Jenkins, J. (2000). *The phonology of English as an international language*. New York: Oxford University Press.

Martin, P. (2003). Winpitch. Pitch Instruments Inc.

Murdock, M. (1987). *Spinning inward*. Boston, MA: Shambhala.

Neville, B. (1989). *Educating psyche: Emotion, imagination and the unconscious in learning.* Victoria: Collins Dove.

Renard, R. (1975). *Introduction to the verbo-tonal method of phonetic correction*, Didier.

Scovel, T. (1969). Foreign accents: Language acquisition and cerebral dominance. *Language Learning, 19*(3,4), 245–254.

Stevick, E. W. (1986). *Images and options in the language classroom*. Cambridge: University Press.

Syntrillium. (2002). *Cooledit 2000*. Syntrillium software.

Tomokiyo, L. M., Le Wang, et al. (2000). *An empirical study of the effectiveness of speech-recognition-based pronunciation tutoring.* Proceedings of ICSLP, Beijing.

Toohey, S. (2000). *Designing courses for higher education*, Buckingham: The Society for Research into Higher Education and Open University.

Troubetzkoy, N. S. (1969). *Principles of phonology (Grundzuge de Phonologie, Travaux du cercle linguistique de Prague).* University of California Press.

Zhang, F., & Newman, D. (2003). *Speech tool.* Canberra: University of Canberra, Australia.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, pp. 377-399, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.4
Web-Based Synchronized Multimedia Lecturing

Kuo-Yu Liu
National Chi-Nan University, Taiwan, R.O.C.

Herng-Yow Chen
National Chi-Nan University, Taiwan, R.O.C.

INTRODUCTION

Over the last decade, the emerging Web technologies have opened a new era for distance education, where online courses can be created and accessed in a very easy way not previously available. Many online courses based on HTML pages thus are now available in cyberspace for synchronous or asynchronous distance learning (Anderson, Beavers, VanDeGrift, & Videon, 2003; Gregory, 1999; Muller & Ottmann, 2000; Shi et al., 2003; Siddiqui & Zubairi, 2000). However, without the support of multimedia, the static HTML pages can only serve as different kinds of simple "dumb" lecture notes on a network. Thus most students may lose interest quickly and eventually give up self-learning (Zimmer, 2003). Furthermore, this kind of unguided, static HTML pages are clearly insufficient for diverse learning needs and for different knowledge domains. With the dramatic development of multimedia technologies, we can integrate various media and provide students with vivid multimedia lectures on the Web. For example, the presentation techniques of online language courses should stress the importance of multimedia (e.g., voice and video) and document interaction flexibility (e.g., random access and repeated play of a specific speech segment) much more than other courses do (Brett, 1998; McLoughlin, Hutchinson, & Koplin, 2002).

The Computer-Assisted Language Learning (CALL) has existed for a long time and has used the computer technology with advanced multimedia and Web technologies to fulfill a certain pedagogical approach (e.g., listening, speaking, reading, and writing) since 1990s (Warschauer, 1996). The purpose of this study is to explore in what ways multimedia and Web technologies can help, and how they can do so in our developed system—the Web-based Synchronized Multimedia Lecture system (WSML)—to make online foreign language teaching and learning more effective

(Chen, Chen, & Hong, 1999). The WSML system has been applied to online language learning that includes English as Second Language (ESL) learning for Chinese students (http://english.csie.ncnu.edu.tw) and enhancement of Chinese for overseas Chinese students (http://chinese.csie.ncnu.edu.tw) in National Chi-Nan University.

The advantages of the WSML system are described as follows:

- **Providing different types of materials can enhance students' capabilities in English/Chinese:** The online teaching materials come from two types of language-learning activities (or sources) with which most students are familiar: (1) lectures with teachers' guidance—the recitation or explanation of an article created and/or used in a real classroom process/experience; and (2) self-learning content without teacher's guidance—the Web/Internet resources beneficial to students' learning. The material involved in a real-classroom activity may include instructors' speech (and/or video), HTML-based lectures, and lecturing events imposed on the HTML lectures. The Web/Internet language material can be HTML-based headline news transcripts and the corresponding news speeches.
- **Providing easy-to-use authoring tools can assist teachers to generate multimedia lectures:** In contrast to the static HTML-based documents, authoring a multimedia document requires much time and work. Therefore, the WSML system provides several authoring tools to assist teachers to create teaching materials, record oral guidance, and capture navigation events without programming skills (Lower, 2001).
- **Web-based multimedia tutoring breaks the limitations of conventional teaching environments:** The online courses provide flexibility for those who are limited by time, distance, or physical ability. Students can choose a suitable course according to their learning situations or suggestions from the teacher. Hence, with the WSML system, students can get adequate practice in listening, reading, and writing.

BACKGROUND

Let's take an English as a Second Language (ESL) course at our university as an example. Instructors usually prepare the teaching materials (e.g., HTML-based reading essays) accessible on the Web for student pre-readings. In class, the instructors use a computer, a microphone, and an LED projector to assist teaching. After opening remarks, the instructors may recite the article to students once at a slower pace before going through other details, such as vocabulary definition and explanation, and so forth. After class, the students are requested to write a reflection on the topic having been taught and to submit the homework through an e-mail or a Web-based submitting interface. Then, instructors receive students' homework, print it out, correct directly on it, and return it back to the students in a later class. This is a model with which we are most familiar and have used for a long while. The computer and network/Web technologies used here are merely to facilitate content exchange between teachers and students.

Imagine that if students could, for example, listen to (and see animated) an online lecture or their own homework as corrected by their teachers, and if a particular text units of the lecture (e.g., keywords and sentences) could dynamically be highlighted in synchronization with the playback of the corresponding speech; students could learn the points more effectively and efficiently. Therefore, the goal of the WSML system is to fulfill the scenarios described above. In what follows, several related works on language learning will be discussed, and the multimedia synchronization issues providing an integrated

synchronized presentation of the WSML system will be investigated.

RELATED WORK

Several language-learning systems have been developed by integrating state-of-the-art multimedia and Web technologies for online language learning. The purpose of these systems is to develop functionalities that support language learning in listening, speaking, reading, and writing skills or online assessment.

Web-CALL (Fujii, Iwata, Hattori, Iijima, & Mizuno, 2000) is an easy-to-use system allowing teachers to add or modify the content of the teaching materials according to their needs. The construction of Web-CALL consists of two units: a Web-page Materials Production Unit (WMPU) and a Learning Support Unit (LSU). The former is a useful tool for teachers to produce Web-page materials without knowledge of programming skills, and the latter enables students to study online language lessons produced by the teacher. Multimedia features such as sound, pictures, and movie files can be attached to a lecture as supplemental materials.

The goal of the Intelligent Web-based Interactive Language Learning (IWiLL) (Kuo et al., 2001) system is to build a networked learning system by integrating language pedagogy, linguistics, computer networks, and multimedia technologies. Two kinds of writing environments in IWiLL are designed to support asynchronous and synchronous writing correction process. The asynchronous writing environment mainly provides functions allowing students to submit essays via the Internet and to examine essays that have been corrected by the teacher. The synchronous writing environment provides functions enabling students and teachers to work on the same essay and to communicate with each other in real-time. In addition, Video-on-Demand technology is also applied in the IWiLL system to support online movie access.

The BRIX system (Sawatpanit, Suthers, & Fleming, 2004) was developed to address the need for a generic language-learning environment that fulfills language-learning activities. Ease of use was important in the design of BRIX, which can yield great benefits to teachers by saving the time and cost of developing courseware. Several functions are implemented for teachers to create teaching materials, such as vocabulary, grammar, discussion, essay, self-test, and quiz. All of the resources can be accessed by the students via the Internet.

The systems described above primarily use the network to distribute and share the language-learning resources. However, multimedia features that are most important for developing online learning systems, particularly for language learning, are less investigated. Web-CALL loosely integrates multimedia files by containing voice information in lectures for listening practice. For the IWiLL system, co-editing and online conversation capabilities are beneficial for essay correction. On the other hand, online movies are other types of teaching materials for listening practice. In contrast to the Web-CALL and IWiLL systems, the BRIX system paid little attention to the use of multimedia materials. It focused on development of functions for fulfilling the instructional activities.

In this article, we present a language-learning system based on developed WSML technology to construct a vivid and vigorous Web-based instruction environment. In the WSML system, teaching materials containing various types of media such as audio-video and navigation events (e.g., pen strokes, highlight, dynamic annotation, tele-pointer, scrolling, etc.) are fully integrated to benefit students in language-learning skills. The elaborate online exercise design is also helpful for self-learning of students. Table 1 shows the summarized assessments of the four language-

Table 1. Summarized assessment of different systems

System / Characteristics	Web-CALL	IWiLL	BRIX	WSML
Support Multimedia teaching materials	medium	medium	low	high
Support Dynamic lecture presentation	no	no	no	yes
Support Online dictionary	no	no	yes	yes
Support Online exercise	yes	yes	yes	yes
Support listening practice	medium	medium	low	high
Support speaking practice	no	no	no	no
Support reading practice	medium	medium	medium	high
Support writing practice	no	high	medium	high

learning systems: the Web-CALL system, the IWiLL system, the BRIX project, and the WSML system, in accordance with their multimedia features and pedagogical approaches.

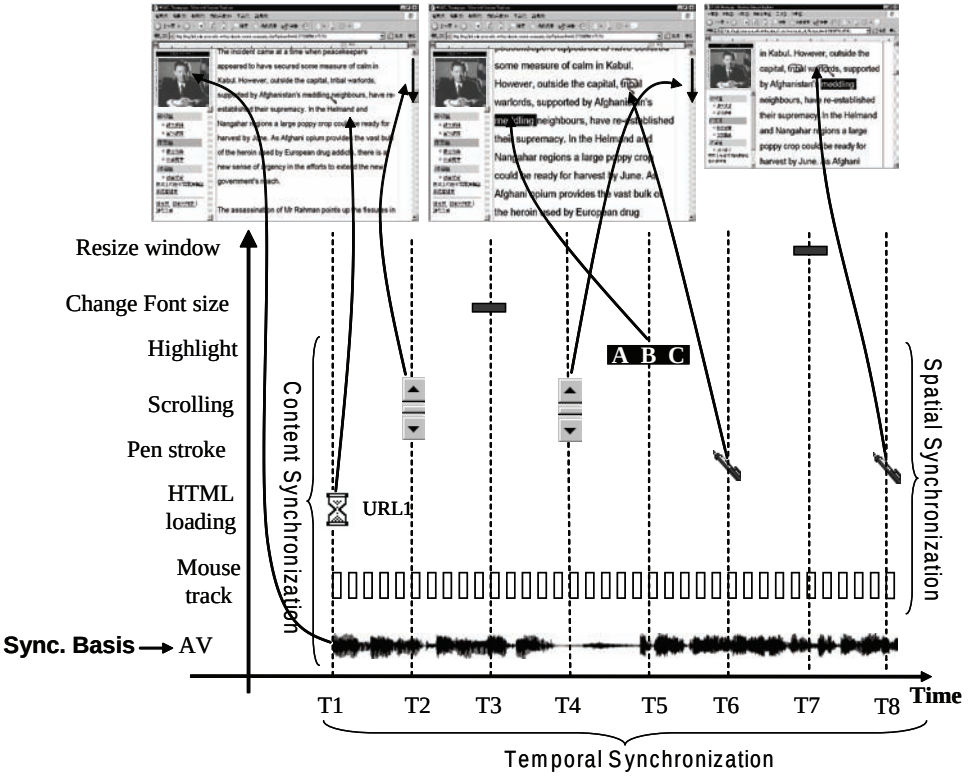
MULTIMEDIA SYNCHRONIZATION

Multimedia systems usually integrate different types of data streams, including continuous media (e.g., audio and video) and discrete media (e.g., text and still image). The synchronization information (i.e., sync-relation) used to render these streams in a correct manner is often interrelated with one another. Thus, multimedia presentation systems must maintain such relations from temporal, spatial, and contextual perspectives and protect them against possible out-of-order and destroyed danger, especially when the streams are going to be presented to end users—the principle of Multimedia Synchronization (Blakowski & Steinmetz, 1996; Effelsberg, Meyer, & Steinmetz, 1993).

In addition, the relations necessary for document presentation may be related to intra-media synchronization (e.g., play audio in continuity without jitter) or inter-media synchronization (e.g., audio-video temporal synchronization in a lip-sync application). In our system, the related media, which include HTML-based document, teacher's voice, and navigation events (e.g., pen strokes, highlight, dynamic annotation, telepointer scrolling, etc.), will be captured during the recording stage. In order to reconstruct the captured process and present the media synchronously, the Adaptive Synchronization Framework (ASF) developed for a better presentation; even the synchronization relations, in particular to the spatial relations, are destroyed by the user interactions/preferences (e.g., font size change, window resize, etc.).

Figure 1 illustrates the handling of synchronization by Adaptive Synchronization Framework. In this scenario, audio is referred to as a presentation time axis (temporal relations) and treated

Figure 1. An example of the integrated synchronized presentation in the WSML system



as a synchronization basis (master stream). The synchronized presentation is described as follows:

T1: The AV and HTML URL1 are loaded at T1. The HTML page is then rendered by the browser and the AV player starts playing the AV lecture at the same time. The mouse-track events would exist all the time during the presentation. Note that the navigation events and the corresponding audio clip should never be presented unless the associated HTML page (namely base page) has been retrieved into the browser, because both navigation-events rendering and audio-clip playing are meaningless if the HTML base slide cannot be loaded successfully (i.e., content synchronization).

T2: A scrolling event is triggered to move down the scroll bar so that the content that is originally out of screen could be displayed. The event is triggered by the teacher during the recording stage.

T3: User enlarges the font size of the document, and the hypertext is subsequently rearranged by the browser.

T4: A scrolling event is triggered to show the content that is out of screen. The event is triggered by the ASF for maintaining content synchronization.

T5: A highlight event over a keyword is invoked at T5.

T6: A pen-stroke event is driven to show a continuous draw effect where the correct coordinates are re-computed by ASF (i.e., spatial synchronization).

T7: At this time, the user narrows the browsing window, and the document layout is changed again.

T8: A pen-stroke event is driven again to show a continuous draw effect. It is rendered according to the correct coordinates re-computed by ASF after the document layout is changed (i.e., spatial synchronization).

IMPLEMENTATION AND FEATURES OF THE WSML SYSTEM

The WSML system integrates audiovisual lectures, HTML slides, and navigation events to provide synchronized presentations. In our environment, teachers use computers to instruct, and a synchronization recorder keeps track of the oral guidance along with several navigation events. The navigation events, such as pen strokes, highlight, dynamic annotation, tele-pointer, and scrolling, are guided media. These media objects and navigated events will be presented dynamically in a browser by using the state-of-the-art, dynamic HTML techniques. In the following, the details about implementation and features of the WSML system will be described. The core of the system includes two perspectives: (1) the synchronized multimedia tutoring and the exercise practice functions for students, and (2) the multimedia lecture-authoring and course management tools for teachers.

For Students

To provide an optimal learning experience for the students, different media objects should be presented synchronously and displayed according to the pre-recorded scenario. In addition, a good exercise system will assist students to have more practices on key points of the lectures.

Voiced Lecture Guidance and Composition Correction

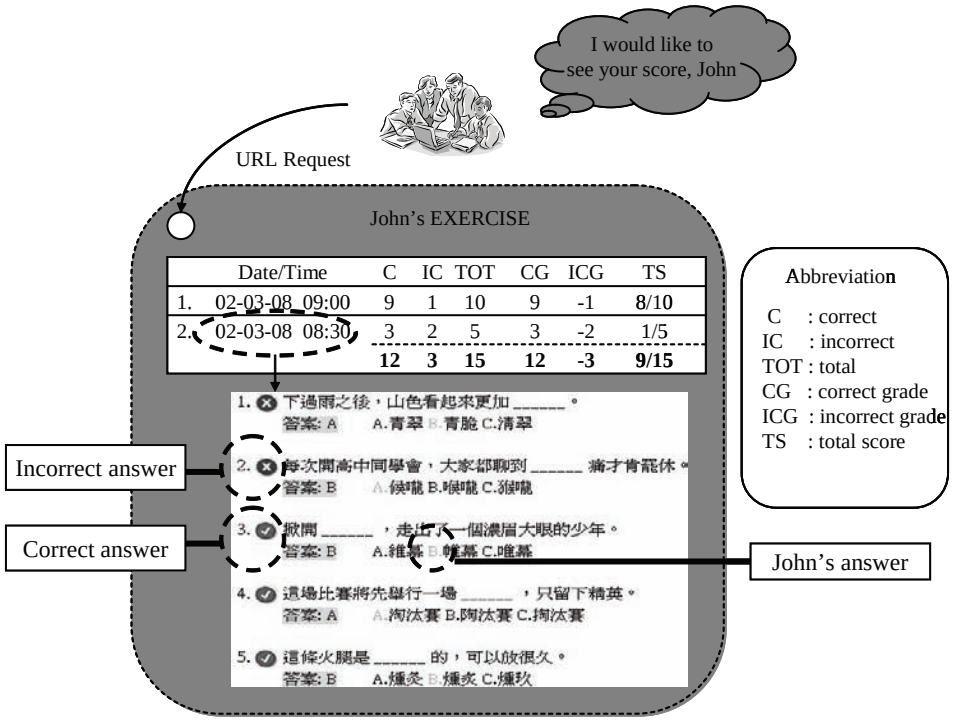
With the authoring tools provided by the WSML system, instructors can record a multimedia document (which may be lecture guidance or a composition correcting results) by navigation events (e.g., mouse moving, annotation, pen drawing, and highlights) and voice, for online access. The function is useful because students can review the multimedia documents again and again. The lecture guidance comprised of recitation with some explanation is a reading enhancement for students. The voice-based composition correction with animated pen-stokes makes students clearly realize what mistakes they made in writing. The function has been applied to online language learning that includes English and Chinese. As Figure 2(a) shows, once a student clicks the play button, he/she can see the composition-correction process completely with the synchronized multimedia presentation. Figure 2(b) shows the voice lecture guidance with tele-pointers and a dynamic dictionary for learning Chinese.

The Online Exercise Function

The online exercise function is designed to evaluate students' learning achievements. Since each student has different degrees of English or Chinese learning, the exercise is categorized into three levels: beginner, intermediate, and advanced level. Three types of exercise—listening exercise, vocabulary testing, and cloze exercise—are provided for students. To a certain extent, each exercise emphasizes different learning goals; for example, listening exercise in intermediate level is different from listening exercise in advanced level. Figure 3 is an example of a Chinese vocabulary test. The following information can be obtained from the exercise function: (1) Students can know how many questions they have answered.

Figure 2. (a) The composition correction with vivid navigation events for learning English; and (b) an example for learning Chinese

Figure 3. The testing results obtaining from the online exercise function



(2) Evaluation results will be given to the student immediately after he or she finishes the test. (3) According to statistical information, students can know what kind of mistakes they made.

FOR TEACHERS

From the teachers' perspective, the WSML system provides several useful multimedia content-authoring tools and a course management tool to create content quickly and manage lectures easily. Moreover, the Web-based exercise management tool helps teachers produce exercise questions more quickly and conveniently.

Multimedia Content Authoring

In the material-authoring stage, this system provides teachers with several tools to facilitate the generation of HTML documents, the recording of oral guidance, and the issuance of some navigation events. These tools include (1) the visualized HTML-editing Tool, and (2) the WSML Recorder.

- **Visualized HTML-editing Tool:** In the WSML system, the most essential teaching media are HTML documents. To enhance the convenience of creating a HTML document, the visualized HTML-editing tool is designed to arrange HTML content visually. Like some commercial editing tools, such as Microsoft FrontPage and Macromedia Dreamweaver, the visualized HTML-editing tool provides "WYSIWYG" features. Teachers can just connect to the WSML Web site and use the online version of the editing tool to generate or manage their HTML documents.
- **The WSML Recorder:** With the prepared HTML documents, teachers can record audiovideo guidance with navigation events by using the WSML Recorder. The recorder consists of three components: a timer, AV encoder, and an event logger. When the teacher starts this recorder, the timer will be initiated to be the timeline of the multimedia document. At the same time, the AV encoder and event logger are also activated to record audio/video data and navigation events, respectively. The recorder detects and stores the synchronization information, including temporal, spatial, and content data, between navigation events and AV lecturing. The navigation events are guided media that are issued by the teacher to emphasize some portions of content or add additional information dynamically. Table 2 shows the collection of navigation events and its description.

Course Management Tool

Having a good course management tool is very important because more and more lectures will be added to the system as time goes on. Additionally, some lectures may need to be updated or modified in the future. For those reasons, we have designed a Web-based course management tool that is described as follows:

(1) Add a new course/lecture: A teacher can manage his or her courses or lectures easily through our system via the Internet. All course information will be recorded in the database (e.g., course, lecture content, teacher's name, etc.). Students can search what they want and need while browsing the lectures. Furthermore, courses and lectures can be classified according to the semesters or course attributes. Students can learn the materials systematically and gradually.

(2) Vocabulary editor: A teacher can first define the vocabularies in the lecture during the recording stage. Through the vocabulary

Table 2. Navigation events and their descriptions

Navigation Event	Description
Audio/Video explanation	Teacher can explain the meaning in a specific sentence and students can practice by imitating the word pronunciation.
HTML Event	Teacher may change the lecture during the instructional process and can add some related reference materials as well as hyperlinks in the lecture.
Scrolling Event	When a HTML document can not be displayed in a viewing window, the scroll bar should be moved to show other portions of content.
Tele-pointer	In the recording stage, teachers often use the mouse to point out the content that they are discussing.
Pen Stroke	The event can simulate strokes of a pen and can be provided for teachers to draw lines or have handwritten effects in the lectures. The result will be presented animatedly in the presentation stage.
Vocabulary Note	A teacher can define the vocabulary appearing in the lecture. The defined vocabularies will become an online dictionary when the students navigate or browse that word. The online dictionary includes the vocabulary explanation, an example, and the pronunciation.
Highlight	The event is useful for teachers to highlight some important words or sentences in the lecture. It is the most frequent feature used to put emphasis on a HTML document.
Dynamic Annotation	To supplement some information, the WSML Recorder also allows teachers to add some textual data dynamically.

editor, the teacher can add the explanation and example sentences of vocabularies and record the pronunciation of vocabularies.

Exercise Management Tool

The exercise management tool provides an easy-to-use authoring interface for exercise creation. The authoring function in the exercise management tool includes the creation of new exercise and re-authoring an exercise. In current stage, several kinds of exercises, such as listening exercise, vocabulary testing, and cloze exercise, have been designed to support specific lecture material. As shown in Figure 4, authoring a cloze exercise is designed to support the understanding in reading a Chinese newspaper.

CONCLUSION

In developing and testing the WSML system for its online use, we consulted many ESL teachers with experience in the Computer Assisted Language Learning (CALL) field. We were delighted to discover that they are happy with the features and existing performance of the WSML system. The easy-to-use lecture-capturing tool and vigorous presentation feature of the WSML made teachers' teaching and homework correction more efficient and effective and more enjoyable, not only for teachers themselves, but for their students. More interesting, multimedia-based lecture presentation increases students' learning motivation and, most important of all, students more clearly understand what mistakes they have made in their homework and why. To this end, the WSML system has been a very helpful online

language tutor in our school. The result of this study is essentially a framework, applicable to different foreign language learning and different teaching scenarios. We hope that the proposed WSML framework will be beneficial for other e-learning applications to explore multimedia relations possibly involved in any teaching activities, and so to develop more effective and useful systems for students' learning.

REFERENCES

Anderson, R., Beavers, J., VanDeGrift, T., & Videon, F. (2003). Videoconferencing and presentation support for synchronous distance learning. *Proceedings of ASEE/IEEE Frontiers in Education*, 2, 13-18, Boulder, CO.

Blakowski, G. & Steinmetz, R. (1996). A media synchronization survey: Reference model, specification, and case studies. *Journal of the IEEE Selected Areas in Communications, 14*(1), 5-35.

Brett, P.A.(1997). Multimedia applications for language learning - What are they and how effective are they? In M. Dangerfield, G. Hambrook, & L. Kostova (Eds.), *East to West* (pp. 171-180). Varna, Bulgaria.

Chen, H.Y., Chen, G.Y., & Hong, J.S. (1999). Design a Web-based synchronized multimedia lecture system for distance education. *Proceedings of IEEE International Conference on Multimedia Computing and Systems (ICMCS99),* June 7-11, Florence, Italy (pp. 887-891).

Effelsberg, W., Meyer, T., & Steinmetz, R. (1993). A taxonomy on multimedia- synchronization. *Proceedings of the Fourth Workshop on Future Trends of Distributed Computing Systems (FTDCS93),* September 22-24, Lisbon, Portugal (pp. 97–103).

Fujii, S., Iwata, J., Hattori, M., Iijima, M., & Mizuno, T. (2000). Web-CALL: A language learning support system using Internet. *Seventh International Conference on Parallel and Distributed Systems Workshops (ICPADS00),* July 4-7, Iwate, Japan (pp. 326-331).

Gregory, D.A. (1999). Classroom 2000: An experiment with the instrumentation of a living educational environment. *Journal of IBM Systems, Special issue on Pervasive Computing, 38*(4), 508-530.

Kuo, C.H., Wible, D., Chen, M.C., Sung, L.C., Tsao, N.L., & Chio, C.L. (2001). Design and implementation of an intelligent Web-based interactive language learning system. *IEEE International Conference on Multimedia and Expo (ICME01),* August 22-25, Tokyo, (pp. 785-788).

Lin, F. & Xie, X. (2001). The practice in the Web-based teaching and learning for three years. *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT01),* August 6-8, Madison, WI (pp. 411-412).

Lower, S. (2001). Systems and software for putting your course on the web. Retrieved April 20, 2004, from Simon Fraser University, Department of Chemistry Web site: *http://www2.sfu.ca/person/lower/cai_articles/WebCAI.html*

McLoughlin, C., Hutchinson, H., & Koplin, M. (2002). Different media for language learning: Does technology add quality? *Proceedings of the Internal Conference on Computers in Education (ICCE 02),* December 3-6, Auckland, NZ (pp. 681-684).

Muller, R. & Ottmann, T. (2000). The 'authoring on the fly' system for automated recording and replay of (tele)presentations. *Journal of Multimedia Systems, 8*(3), 158-176.

Sawatpanit, M., Suthers, D., & Fleming, S. (2004). BRIX: Meeting the requirements for online second

language learning. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS04),* January 5-8, Big Island, Hawaii, (pp. 4-13).

Shi, Y., Xie, W., Xu, G., Shi, R., Chen, E., Mao, Y., & Liu, F. (2003). The smart classroom: Merging technologies for seamless tele-education. *IEEE Pervasive Computing Magazine, 2*(2), 47-55.

Siddiqui, K.J. & Zubairi, J.A. (2000). Distance learning using Web-based multimedia environment. *Proceedings of Academia/Industry Working Conference (AIWORC00),* April 27-29, Buffalo, New York, (pp. 325-330).

Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia language teaching* (pp. 3-20). Tokyo: Logos International.

Zimmer, J. E. (2003). Teaching effectively with multimedia. Visionlearning. Retrieved April 16, 2004, from *http://www.visionlearning.com/library/module_viewer.php?mid=87*

KEY TERMS

Adaptive Synchronization Framework (ASF): A framework that adopts a mechanism to re-compute the synchronization relations that are destroyed by unexpected events (e.g., window resizing or font size change), so as to restore a presentation system to synchronization.

AV Encoder: An encoder used to encode the input audio/video signal to specific digital format.

Computer Assisted Language Learning (CALL): The topics discuss the teaching practices and research related to the use of computers in the language classroom.

Event Logger: A text file used to record the timestamps, attributes, and types of navigation events invoked by the teacher during the recording stage. The log is treated as synchronization information for dynamic presentation.

Multimedia Synchronization: Multimedia systems usually integrate different types of data streams, including continuous media (e.g., audio and video) and discrete media (e.g., text and still images). The information (e.g., temporal, spatial, and content relations) for displaying these data streams is often interrelated. Thus multimedia systems must guarantee such relationships between streams against partial loss or being nonfunctional when the streams are transmitted after capture and presented to end users.

Navigation Events: The events, such as pen stroke, tele-pointer, document scrolling, annotation, highlights, and the like, invoked by the teacher during the recording stage will be captured as guided media and thus can be presented dynamically to the end users.

Timer: The timer is used to determine timestamp of navigation events invoked by the teacher during the recording stage.

This work was previously published in Encyclopedia of Distance Learning, Vol. 4, edited by C. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers, and G.A. Berg, pp. 2019-2028, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.5
Teaching, Learning and Multimedia

Loreen Marie Butcher-Powell
Bloomsburg University of Pennsylvania, USA

ABSTRACT

"We must not forget that almost all teaching is Multimedia" (Schramm, p.37). Today, the magnetism of multimedia is clearly oblivious via the use of streaming video, audio clips, and the Internet. Research has shown that the use of multimedia can aid in the comprehension and retention of student learning (Cronin & Myers, 1997; Large Behesti, Breulex & Renaud, 1996; Tennenbaum, 1998). As a result, more educators are utilizing Web-based multimedia materials to augment instruction online and in the classroom. This chapter provides a theoretical framework for transforming Student Centered Discussion (SCD), a traditional based pedagogy strategy, to a new multimedia pedagogy SCD strategy. The new multimedia SCD pedagogy represents a new way of teaching and learning. As a result, positive responses and feedback have been collected from students in their ability to interpret facts, compare and contract material, and make inferences based on recall of information previously presented or assigned in article readings.

INTRODUCTION

Research has shown that students can integrate information from various sensory modalities into a meaningful experience. For example, students often associate the sound of thunder with the visual image of lightning in the sky. When the cognitive impact of two given interaction modalities differ enough, different learning modes can be induced. Moreover, an interaction modality, which affects a learning mode, also has consequences for the learning performance (Guttormsen, 1996, 1997). Therefore, a teacher is faced with the need to integrate various combinations of sensory modalities, such as text, still images, motion, audio, animation, etc., to promote the learning experience.

Multimedia is multisensory; it engages the senses of the students. Multimedia can be defined in a variety of ways, but in this chapter, the term "multimedia" refers to a Web-based interactive computer-mediated application that includes various combinations of text, sound, still images, audio, video, and graphics. Multimedia is also interactive; it enables both the student and the

teacher to control the content flow of information (Vaughan, 1998). A major part of using multimedia in instruction involves engaging students in sense-making activities, such as conversations and chats about external representations that use concepts, symbols, models, and relationships. As a result, multimedia has introduced important changes in the educational system and has impacted the way teachers communicate information to the student (Neo & Neo, 2000).

Learning

Learning is fundamentally built up through conversations between persons or among groups, involving the creation and interpretation of communication (Gay & Lentini, 1995; Schegloff & Sacks, 1973; Schegloff, 1991). More importantly, learning is established and negotiated through successive turns of action and conversations (Gay et al., 1995; Goodwin & Hertage, 1986; Schegloff, 1991). Thus, conversations are means by which people collaboratively construct beliefs and meanings as well as state their differences.

Brown, Collins, and Duguid (1989) argued that learning involves making sense of experience, thought, or phenomenon in context. They hypothesized that student representation or understanding of a concept is not abstract and self-sufficient, but rather it is constructed from the social and physical context in which the concept is found and used. Further, Brown et al. (1989) emphasized the importance of implicit knowledge in developing understanding rather than acquiring formal concepts. It is, therefore, essential to provide students with authentic experiences with the concept.

Students can engage in learning conversations in distributed multimedia environments. Multimedia technologies, such as graphics, simulations, video, sound, and text, allow instructors to use multiple modes and representations to construct new understanding and conceptual change of enhancing student knowledge. Brown et al. (1989) stated that learning involves making sense of thoughts, experiences, or phenomena in contexts. Multimedia allows for the accommodation of diverse learning styles. Different media provide different opportunities for communication and activities among students. For example, online conversations provide a common background or mutual knowledge about beliefs and assumptions during conversation.

The Distinct Ways of Learning

There are multiple ways of learning. Four of the most common and distinct ways to learn are independent learning, individual learning, cooperative group learning, and collaborative group learning (Kawachi, 2003). For the purpose of this chapter, it is important to understand the differences between cooperative and collaborative learning.

Traditionally in a cooperative learning environment, knowledge is learned by the student via the teacher or other students repeating, reiterating, recapitulating, paraphrasing, summarizing, reorganizing, or explaining the concepts. Meanwhile, in collaborative learning, knowledge is not learned by the student via the teacher, but rather knowledge is learned via an active dialogue among students who seek to understand and apply concepts. Using multimedia in collaborative environments allows students to participate in genuine learning activities by which they can reflect as well as modify their understanding of concepts (Brown et al., 1989; Gay, Sturgill, Martin, & Huttenlocher, 1999; Harasim, Hiltz, Teles, & Turoff, 1995; Wegerif, 1998; Murphy, Drabier, & Epps, 1997). The ability to read and respond to a message posted to an online forum creates opportunities for the creation of knowledge.

With the use of multimedia, students can utilize the information presented to them by the teacher, and represent it in a more meaningful way, using different media elements. Fortunately, there are many multimedia technologies that are available for teachers to use to create innovative

and interactive courses. A review of literature on multimedia educational tools revealed some interesting innovative and rich multimedia-based learning tools. Jesshope, Heinrich, and Kinshuk (n.d.) researched the ProgramLive application. The ProgramLive application is a rich multimedia-based tutorial of the Java programming language. ProgramLive's interface represents a notebook, within a browser. There are tabs to the side of the notebook display that can be used for navigation of the material, as well as pop-up explanations of key concepts.

Millard (1999) developed an Interactive Learning Module for Electrical Engineering Education and Training titled the Interactive Learning Modules (ILM). ILM presents Web-based multimedia tutorials created with Macromedia Director. ILM provides a mechanism for the creation of supplementary material for lectures and collaborative problem solving and simulation environments. ILM is highly modular for the usage of various materials to be used in multiple courses. Similarly, the Multimedia Learning Environment (MLE) developed by Roccetti and Salomoni (2001) is a networked educational application that also provides course material in a student-based manner. MLE provides a virtual learning environment, through a client application, where the multimedia educational material is structured in Adaptive Hypermedia, from which sets of hypermedia pages are dynamically retrieved and presented to the student via tailoring the contents and presentation style to the students needs (Roccetti et al., 2001).

Further, Jesshope, Heinrich, and Kinshuk (n.d.) are currently developing an integrated system for Web-based education called the Teaching Integrated Learning Environment (TILE). This system uses Web-based delivery of course material, including interactive multimedia presentations, problem solving, and simulation environments in which students learn by doing. Like MLE, TILE provides students with an interactive multimedia environment and instructors with a multimedia environment for managing, authoring, monitoring, and evaluating student learning.

The multimedia educational tools, described above, have been traditionally used in two ways, either as a vehicle for students to learn theory and application beyond the subject matter or as a tool used by the teacher to support teaching. As a result of multimedia educational tools, teachers are faced with a significant need to provide a more multimedia-based approach to learning, and to create a new educational pedagogy that emphasizes collaborative learning via multimedia.

Numerous studies have been conducted in the attempt to determine how effective multimedia is in teaching (Blank, Pottenger, Kessler, Roy, Gevry, Heigel, Sahasrabudhe, & Wang, 2002), however, very few studies have been conducted to illustrate and determine the factors that may aid in a new multimedia pedagogy strategy for teaching. This chapter was designed to provide the theoretical framework for how teaching is enhanced using Web-based multimedia. The objective of the chapter will be to explain the latest pedagogical teaching strategies for utilizing interactive Web-based multimedia educational tools. As a result, this chapter will provide instructors with a positive and effective example for utilizing Web-based multimedia in teaching.

A NEW GLOBAL ENVIRONMENT FOR LEARNING

A new multimedia pedagogical model for learning in a collaborative environment was incorporated in the Information Science and Technology (IST) undergraduate program at Pennsylvania State University (PSU), Hazleton, Pennsylvania (USA). The transformation from a traditional lecture-based model to an interactive Web-based multimedia application was accomplished using A New Global Environment for Learning (ANGEL). ANGEL is PSU's course management system (CMS) that is currently in use within the University's system

(Pennsylvania State University, n.d.). ANGEL is an interactive Web-based multimedia application developed by CyberLearning Labs Incorporated, for constructive, collaborative, inquiry-based, and problem-solving Web-based learning. ANGEL allows for interaction, testing, presentations, audio, video, forums, file submissions, and many more multimedia features. Through ANGEL, instructors are able to use multimedia effectively in aiding in the students' learning and retention process (Pennsylvania State University, n.d.).

Pedagogical Strategy

Beginning with the philosophy that "learning is not a spectator sport" (Chickering & Gomson, n.d.), students are encouraged to get involved in their educational experience. The probability of students' learning improving by getting involved, talking and writing about what they have learned, relating it to past experiences, and most importantly applying it to their daily lives, is much greater than by students sitting in classes listening to teachers, memorizing prepackaged assignments, and spitting out answers. The goal of each course is to provide the students with a challenging, critical thinking, novel, technology-focused, and learner-centered educational experience, where they learn by pursuing knowledge, improving basic communication skills, and, most importantly, taking responsibility for their own learning (Brown et al., 1989).

To obtain such a goal, the following procedures were used. The classes were structured toward creating a problem-based learning (PBL) and a student-centered discussion (SCD) environment for students utilizing a multimedia course management system, ANGEL. PBL is traditionally used in courses that provide more student-centered learning experiences. The origins of PBL began in the medical education field (Barrows, 1986, 1999). PBL is a student-centered pedagogical approach in which learning is taught through suggested real-world problems. PBL establishes the importance of clearly formulated effective real-world problems. An effective problem has a realistic context and is couched in appropriate vocabulary. The problem should be complex and ill-structured, without clear-cut, easy answers or nuances and subtitles that are not immediately apparent. Moreover, the problem should support both discovery and self-directed learning while engaging the interest and the curiosity of the student (Desmarchais, 1999).

SCD is a delivery system for the application of educational goals in the classroom. This process is accomplished by integrating basic discussion skills in the classroom. The technique is one that models after all competency levels of Blooms Taxonomy in a limited structured time period. The discussion and team-building process that occurs in SCD promotes the active engagement of the students in their own educations. This technique requires the student to actively take responsibility for conducting a productive and meaningful discussion (Wright & Shout, n.d.). SCD has been proven to be an interactive model that encourages students to develop effective communication and interpersonal skills as well as strengthen critical-thinking skills (Butcher-Powell & Brazon, 2003). Moreover, Wright et al. (n.d.) stated that this model is effective regardless of discipline or knowledge base.

The addition of multimedia technology into a PBL and a SCD environment further enhances the students' learning experience. Figure 1 illustrates this focus.

Class Structure

The original version of the PBL and SCD was modified in order to incorporate multimedia into the class. The classes meet every Tuesday and Thursday for approximately one hour and 10 minutes for one 15-week and one 14-week semester. On every Tuesday, the PBL and SCD models were used in class. To accomplish a cohesive learning experience, the students are first divided into small

groups consisting of four students. Each group has 30 minutes to discuss, ponder, debate, question, learn, and solve the problem from a video clip stored on ANGEL. Additional resources, such as articles, notes, and existing Web cases were made available to the students in a digital library on ANGEL. After the 30 minutes were over, the entire class was combined into one big circle, upon which students received another video clip that expanded upon the first video clip. The students were to engage in a larger PBL and SCD environment to create a solution to the problem identified in the video clips.

After the class session was over, students were required to log on to ANGEL and write a brief summary of what they learned from the video clips, and how that relates to their lives. In addition, the students were required to search the Web and find an article on the subject matter and post it to ANGEL for Thursday's class. On Thursday, an interactive PowerPoint lecture supporting Tuesday's video clip and the students' PBL and SCDs were presented to the class and remained available in ANGEL for later student usage. After the 30 minute lecture, the remaining 40 minutes were left for the students to find and discuss an article or Web site in the forum section of ANGEL. Moreover, each student was also required to read and elaborate upon at least one other student's summary of a related article or Web site.

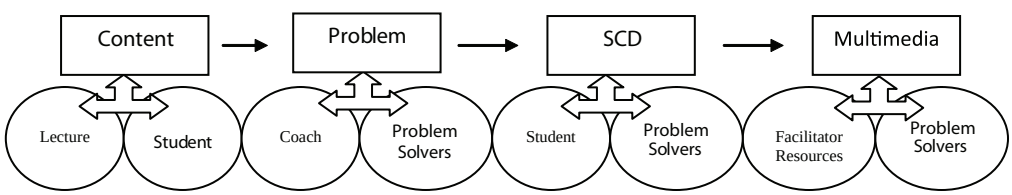
After the material was taught utilizing the defined methods, the students were required to take an online interactive time test via ANGEL. Traditionally, the test consisted of random multiple-choice, true and false, matching, and short-answer questions. ANGEL allowed authenticated users, during a specified time frame, various combinations of questions, and timed the test. More importantly, ANGEL allowed the teacher to make prerecorded video available during the test in order to guide a student through each excursion of the test. And, finally, ANGEL allowed the teacher to pregrade the test so that the students had immediate access to their grades after their tests were submitted. However, the answers to the test were only available for review after all students completed the test.

As illustrated above, this new multimedia pedagogical strategy allowed for inquiry to be accomplished via text mining and visualization tools. As a result, students were able to explore the emerging problems, solutions, and trends in their fields of study. Moreover, collaborative learning was also achieved via live links and remote-controlled "video" sessions by reviewing multimedia recorded lecture sessions that encouraged students to interact with instructors and digital libraries.

Data Source and Collection Instrument

The data population consisted of undergraduate students (n = 32) in the Information Science and Technology major at PSU Hazleton, Pennsylvania, USA. The students ranged in age from 20 to 28 (M = 20.51, SD = 2.68). Academic classification was represented by juniors (n = 26) and seniors (n = 8). The teaching team consisted of one instructor utilizing two classes to explore student

Figure 1: The multimedia PBL–SCD curriculum model



perceptions and experiences with the cohesive multimedia pedagogical strategy. Data were gathered in pre- and postquestionnaires and in student diaries. The pre- and postquestionnaires consisted of 11 questions to assess the students' interest in group project work and whether or not they were motivated in their project development. The questionnaire also tried to determine the students' levels of understanding and critical-thinking skills. The questionnaire was measured using a five-point Likert scale. The scale measurements were 1, strongly agree; 2, agree; 3, undecided; 4, disagree; and 5, strongly disagree. The scales and all questions in the questionnaire were developed after a review of literature guided by the theoretical base. An expert panel of faculty and doctoral-level graduate students evaluated the face and content validity of the questionnaire.

RESULTS AND DISCUSSION

The feedback at the time of writing indicates that the new multimedia pedagogy approach used in the classes is fully endorsed by students and leads to intense student engagement and an effective level of learning. Thus, over 80% (87.7%) of the students ($n = 32$) had a computer with Internet access at home and were able to fully take advantage of the new multimedia way of learning. At the beginning of the course, most (90.6%) of the students were excited and eager to be taught using the new multimedia pedagogy. Furthermore, almost all of the students (96.87%) at the end of the course felt they had learned a great deal of theory, content, and application with the new multimedia pedagogy. Consistently, students (96.87%) indicated that they would like to see this new multimedia pedagogy applied to all of their undergraduate classes. The perceived benefit of the new multimedia pedagogy to the student mean was 1.92 ($SD = 0.63$). This indicated that students agreed with the perceived benefits of the new multimedia pedagogy strategy.

Inferential *t*-tests were used to determine if significant differences ($p < 0.05$) existed in the perception of additional workload as a result of the adoption of the new multimedia strategy. The *t*-test revealed that no significant differences ($t = 1.59$, $p = 0.114$) existed in the additional workload mean between the students who liked and did not like the new multimedia way of learning. Furthermore, no significant differences existed in the additional workload mean between the students who earned a B or better and the students who earned a B- or less. More importantly, no significant differences existed between the students who earned a B or better or the students who earned a B- or lower, as they strongly agreed that this was a great way to learn.

Moreover, the students' experiences with this new multimedia pedagogical strategy were also demonstrated throughout the semester by the student's ability to interpret facts, compare and contrast material, and make inferences based on recall of the information previously presented or assigned information in the given article. Most importantly, the student's readiness for the class was instantly recognized.

Overall, the new multimedia pedagogical strategy represents a new way to develop and deliver classroom discussions and material. In traditional classroom discussions, the faculty member assumes the role of the discussion leader. He or she identifies the questions that will be discussed and maintains a teacher-centered structure. In contrast with traditional classroom discussions, the new multimedia pedagogical strategy online approach is much more student-centered through the usage of multimedia enhancements. As a result, the instructor assumes the role of discussion facilitator, instead of discussion leader. This shift from discussion leader to discussion facilitator forces students to become responsible for their behavior and interaction in the discussion.

Limitation

The integration of multimedia into the classroom and online is essential if multimedia is to become a truly effective educational resource. However, the integration and the change in pedagogy are difficult, time consuming, and resource-intensive tasks. Research has shown that teachers need time working with the technology before they will be at a level of comfort to change or modify their pedagogy (Redmann et al., 2003).

CONCLUSIONS

Despite a rapidly growing recognition of the potential impacts of such interactive multimedia developments, teachers have relatively little understanding of the extensive benefits surrounding the use of new multimedia resources in the classroom. The dearth of knowledge exists in a wide variety of domains, including but not limited to the design and development of new systems and tools to retrieve and manipulate documents, as well as the uses and impacts of such new tools on both learning and problem solving. Previous research documents the improvement in learning as a result of the use of multimedia (Blank et al., 2002), while other researchers have not found any significant differences in learning between multimedia-based and traditional-based pedagogical approaches (Moore & Kearley, 1996). As efforts to illustrate the impact of multimedia in instruction continue, one fact remains: Modern multimedia-based pedagogical approaches have roots of the oldest traditional human communication methods. However, in the future, multimedia pedagogical strategies will have a much more profound impact on how instructors approach and engage students in the process of education and communication.

As the complexity of using multimedia in the classroom continues, new teacher and student technologies will become robust, and ad hoc methodologies will give way to more interactive student competence for learning.

REFERENCES

Barrows, H. (1999). The minimum essentials for problem-based learning. Retrieved March 2003 from the World Wide Web: http://www.pbli.org/pbl/pbl_essentials.htm

Barrows, H. S. (1986). A taxonomy of problem based learning methods. *Medical Education, 20*, 481–486.

Blank, G. D., Pottenger, W. M., Kessler, G. D., Roy, S., Gevry, D., Heigel, J., Sahasrabudhe, S., & Wang, Q. (June, 2002). Design and evaluation of multimedia to teach java and object oriented software engineering. *Proceedings of the American Society for Engineering Education*. Montreal, Canada.

Brown, J. S., Collins, A., & Dugid, P. (1989). Situated cognition and the Culture of Learning. *Educational Researcher, 18*, 32–42.

Butcher-Powell, L. M., & Brazon, B. (February, 2003). Workshop: Developing interactive competence: Student centered discussion. *Journal of Computing Science in Colleges, Proceedings of the 18th Annual CCSC Eastern Conference, Bloomsburg, PA, 18*(3), 235–240.

Chickering & Gomson. (n.d.). Seven principles of good practice in undergraduate education. Retrieved April 2003 from the World Wide Web: http://www.hcc.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/7princip.htm

Cronin, M. W., & Myers, S. L. (Spring 1997). Effects of visual versus no visuals on learning outcomes from interactive multimedia instructions. *Journal of Computing in Higher Education, 8*(2), 46–71.

Desmarchais, J. E. (1999). A Delphi technique to identify and evaluate criteria for construction of PBL problems. *Medical Education, 33*(7), 504–508.

Gay, G., & Lentini, M. (1995). Use of communication resources in a networked collaborative design environment. *Journal of Computer-Mediated Communication,1*(1). Retrieved April 2003 from the World Wide Web: http://www.ascusc.org/jcmc/vol1/issue1/IMG_JCMC/ResourceUse.html

Gay, G., Sturgill, A., Martin, W., & Huttenlocher, D. (1999). Document-centered peer collaborations: An exploration of the educational uses of networked communication technologies. *Journal of Computer-Mediated Communication, 4*(3). Retrieved January 13, 2004 from: http://www.ascusc.org/jcmc/vol4/issue3/gay.html

Goodwin, C., & Hertage, J. (1986). Conversation analysis. *Annual Review of Anthropology, 19,* 283–307.

Guttormsen Schar, S. G. (1996). The influence of the user-interface on solving well- and ill-defined problems. *International Journal of Human–Computer Studies, 44,* 1–18.

Guttormsen Schar, S. G. (1997). The history as a cognitive tool for navigation in a hypertext system. In M. J. Smith, G. Salvendy, & R. J. Koubek (Eds.), Vol. 21B, pp. 743–746.

Harasim, L., Hiltz, S. R., Teles, L., & Turoff, M. (1995). *Learning networks: A filed guide to teaching and learning online.* Cambridge, MA: MIT Press.

Jesshope, C., Heinrich, E., & Kinshuk. (n.d.). *Online education using technology Integrated Learning Environments.* Massey University, New Zealand. Retrieved February 2003 from the World Wide Web: http://www.tile.massey.ac.nz/publicns.html

Kawachi, P. (2003). Choosing the appropriate media to support the learning process. *Media and Technology for Human Resource Development, 14*(1&2), 1–18.

Large, A., Behesgti, J., Breuleux, A., & Renaud, A. (1996). Effect to animation in enhancing descriptive and procedural texts in a multimedia environment. *Journal of the American Society of Information Science, 47*(6), 437–448.

Millard, D. M. (1999). Learning modules for electrical engineering education and training. *Proceedings of the American Society for Engineering Education.*

Moore, M.G. & Kearsley, G. (1996). *Distance Education: A Systems View.* Wadsworth Publishing.

Murphy, K. L., Drabier, R., & Epps, M. L. (1997). Incorporating computer conferencing into university courses. *1997 Conference Proceedings: Fourth Annual National Distance Education Conference* (pp. 147–155). College Station, TX, USA: Texas A & M University. Retrieved January 2003 from the World Wide Web: http://disted.tamu.edu/~kmurphy/dec97paphtm

Neo, M., & Neo, T. K. (2000). Multimedia learning: Using multimedia as a platform for instruction and learning in higher education. *Paper presented at the Multimedia University International Symposium on Information and Communication.*

Pennsylvania State University. (n.d.). Overview and tools. Retrieved February 9, 2003 from the World Wide Web: http://cms.psu.edu

Redmann, H. D., Kotrlik, W. J., & Douglas, B. B. (2003). A comparison of business and marketing teachers on their adoption of technology for use in instruction: Barriers, training, and the availability of technology. *NABTE Review, 30,* 29–35.

Roccetti, M., & Salomoni, P. (2001). A Web-based synchronized multimedia system for distance edu-

cation. *Proceedings of the 16th ACM Symposium on Applied Computing* (pp. 94–98).

Scardamalia, M., & Bereiter, C. (1993). Collaborative knowledge building. In E. DeCorte, M. C. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 41–66). Berlin: Springer-Verlag.

Schegloff, E. A. (1991). Conversation analysis and socially shared cognition. In L. Resnick, J. Levine, & S. D. Bernard (Eds.), *Socially shared cognition* (pp. 150–172). Washington, DC: American Psychological Association.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 7*, 289–327.

Schramm, W. (1977). *Big media, little media*. Beverly Hills, CA: Sage Publications.

Tennenbaum, R. S. (1999). *Theoretical foundation of multimedia*. New York, NY: Computer Science Press.

Vaugh, T. (1998). *Multimedia: Making it work* (4th ed.). Berkeley, CA: Osborne/McGraw Hill.

Wegerif, R. (1998). The social dimension of asynchronous learning networks. *Journal of Asynchronous Learning Networks, 2*(1), 34–39.

Wright, D., & Shout, L. (n.d.). Developing interactive competence through student-centered discussion. Retrieved March 2003 from the World Wide Web: http://home.kiski.net/~dwright/scd/hme.html

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, pp. 60-72, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.6
Interactive Multimedia for Learning and Performance

Ashok Banerji
Monisha Electronic Education Trust, India

Glenda Rose Scales
Virginia Tech, USA

ABSTRACT

Developments in information and communication technologies (ICT) are rapidly transforming our work environments and methods. Amongst these changes, the advent of interactive multimedia technology has meant new approaches to instruction, information and performance support implementations. The available resources can be amalgamated in a suitable way to create an enabling environment for learning, training and performing. Concise descriptions of the salient aspects are presented along with basic design principles for communication and performance support. Guidelines for design and suggestions for implementation are provided for the benefit of the practitioners.

INTRODUCTION

Undoubtedly, the advent of computers and communication technology has forever changed our daily lives. Today, we have the fantasy amplifiers (computers), the intellectual tool kits (software and hardware), and the interactive electronic communities facilitated by the Internet that have the potential to change the way we think, learn, and communicate. However, these are only tools. The late Turing Award winner Edsger Dijkstra said, "In their capacity as a tool, computers will be but a ripple on the surface of our culture. In their capacity as intellectual challenge, they are without precedent in the cultural history of mankind" (Boyer et al., 2002). The onus is on us, our innovative ideas as to how we harness the technology for education, training, and business

in order to lead or lag in the new social order. In this regard, we may remember that Charles Darwin said, "It's not the strongest of the species who survive, nor the most intelligent, but the ones most responsive to change."

In this chapter, we will review these current developments in teaching and learning from a broader performance support systems perspective. Then we will suggest a performance-centered design approach in support of developing teaching and learning solutions for the knowledge worker of today.

Lessons from the Past

There are many examples from the past indicating the rush to implement cutting-edge technologies (Marino, 2001). All of these began with a grand promise as a total solution to a long-standing problem. For example, in 1922 Thomas Edison predicted that "the motion picture is destined to revolutionize" the educational system and will largely supplement textbooks. Radio was hailed with the promise to "bring the world to the classroom." Similarly, educational television was touted as a way to create a "continental classroom" (Cuban, 1986). How much of these hopes have been met as of today?

On similar lines, recently, there has been much hype about interactive multimedia and the Internet as the remedies for all problems in training and education. However, as a knowledge resource, multimedia productions, the Internet, and a library have similar attributes. It is particularly wrong to assume that putting all the information on the Internet will make learning happen. The Internet is useful, but it does not guarantee learning any more than a good library ensures creating knowledgeable persons (Clark, 1983).

From a technocratic perspective, there is a tendency to assume that installing computers and networks will solve every conceivable problem. However, the value and benefits of technology will come only through leveraging it for dynamic and strategic purposes that place the focus first on learning and performing and second on the technology (Dede, 1998; Bare & Meek, 1998).

The key lessons from the past indicate that including performance-centered design techniques tends to improve the usability of the information or learning systems. As we move from the "Information Age" into the "Knowledge age," it is important to consider technological solutions to support teaching and learning (Reeves, 1998). In the transition from the "Old Economy" to the "New Economy," a key outcome of the transformation is a dramatic shift from investments in physical capital to investments in human or intellectual capital. A well-designed holistic approach toward training and development is therefore needed to support the learning needs of the knowledge worker (McArthur, 2000). In this regard, we need to consider the benefits of a user- and performance-centered approach from the standpoint of design. The remaining portion of this chapter will discuss how the electronic performance support systems approach can help in the challenges associated with the new paradigm.

PERFORMANCE SUPPORT SYSTEMS

There are three primary impacts of information and communication technologies (ICT). These are the methods in which the following occur:

(a) Information is distributed and retrieved.
(b) Knowledge and expertise are stored and acquired.
(c) Skills are learned and transferred.

These technologies have made important impacts in transforming education, training, and skill development approaches. In 1991, Gloria Gery introduced a framework for electronic support (Gery, 1991, 2002). While definitions vary, it is widely agreed that performance support systems do the following:

- Enable people to perform tasks quickly, because they provide integrated task structuring, data, knowledge, and tools at the time of need
- Do not tax the performer's memory or require performers to manipulate too many variables
- Enable task completion, with learning as a secondary consequence

Taking a broader view, we can say that an electronic performance support involves "a human activity system that is able to manipulate large amounts of task related information in order to provide both a problem solving capability as well as learning opportunities to augment human performance in a job task" (Banerji, 1999a). Such systems provide information and concepts in either a linear or a nonlinear way, as they are required by a user. The EPSS concept provides a holistic design framework encompassing a custom-built interactive guidance, learning, information, and knowledge support facility that is integrated into a normal working environment. Such systems are concerned with effective human–task interaction in which the computer provides an interface to various job tasks and becomes an aid in achieving efficient task performance.

Components and Types of EPSS

In most modern workplaces, computers are used for decision making, task performance, task sequencing, planning, and also learning, thereby replacing many manual methods. In such situations, the work is not done solely by people or solely by computers but by human-computer systems. The computers and communication technology thus act as a powerful tool by providing an interface to the basic job tasks that are involved. People and computers thus tend to work cooperatively and symbiotically, combining the advantages of the powers of each in order to achieve more effective job performance (Licklider, 1960).

Thus, human-task interaction within the human activity system (HAS) forms the foundation of EPSS. The HAS involves the following three subsystems, as shown in Figure 1:

(a) The tool subsystem
(b) The task subsystems
(c) The people subsystem

The tool subsystem provides an interface to various job tasks and becomes an effective aid in achieving efficient task performance. It can also be a means for improving performance. However, the performance generally gets hindered in the absence of an appropriate "interface." These are the barriers of task performance that a support system should strive to minimize. The dimensions of these barriers include knowledge, skill, information, decision, processes, and procedures. The function of an EPSS would be to reduce the "permeability" of the interface through appropriate means. These include eLearning facility and Knowledge Management, among many others (Dickelman & Banerji, 1999).

There could be three principal ways in which the "tools system" interfaces the "task system," and, three broad classifications of EPSS can be made depending on how they render support in task performance:

Type 1: In this type, tasks are performed with computer and software tools, such as word processors, spreadsheets, and so on. Support for this type of application is tied with the software tools and, therefore, can be called *software-integrated EPSS*. The simplest examples are cue cards, animated help in Microsoft applications (Figure 2), and wizards.

Type 2: In this type, computer-based tools mediate the organizational tasks and practices, such as banking systems, enterprise resource planning systems, air ticket booking, along with hotel and car booking systems, and so

Figure 1: Concept map of human activity system

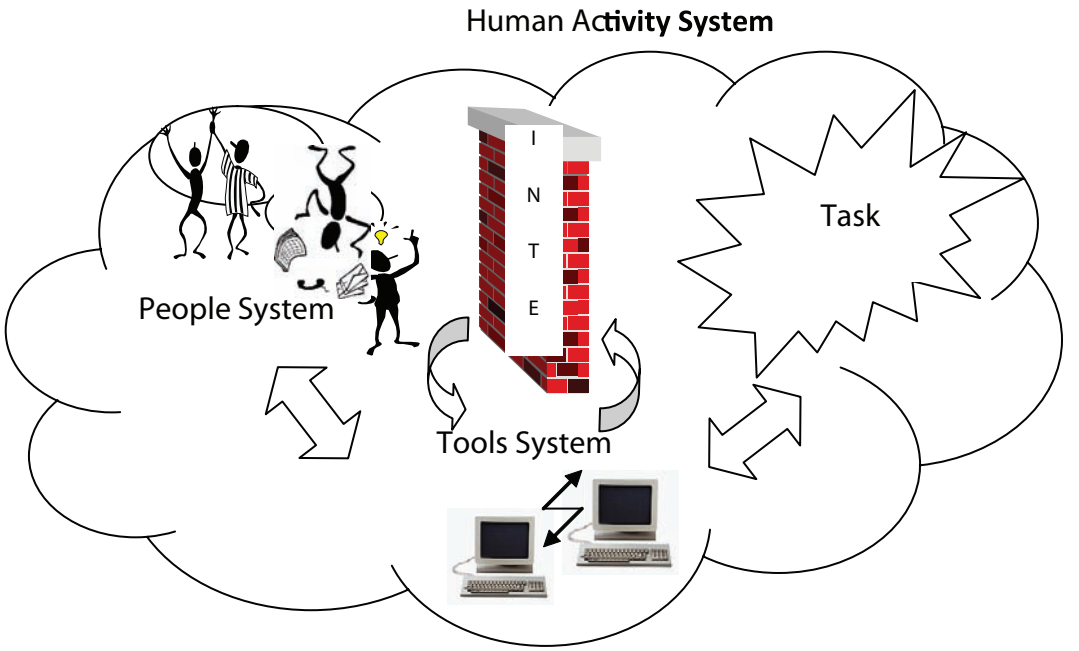
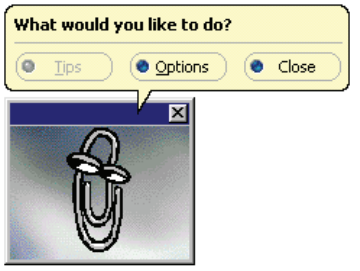


Figure 2: Animated help in Microsoft applications



on. Supports are needed as an integrated part of this type of application so that the user can perform competently with minimal training. These types of applications can be called *job-integrated EPSS.*

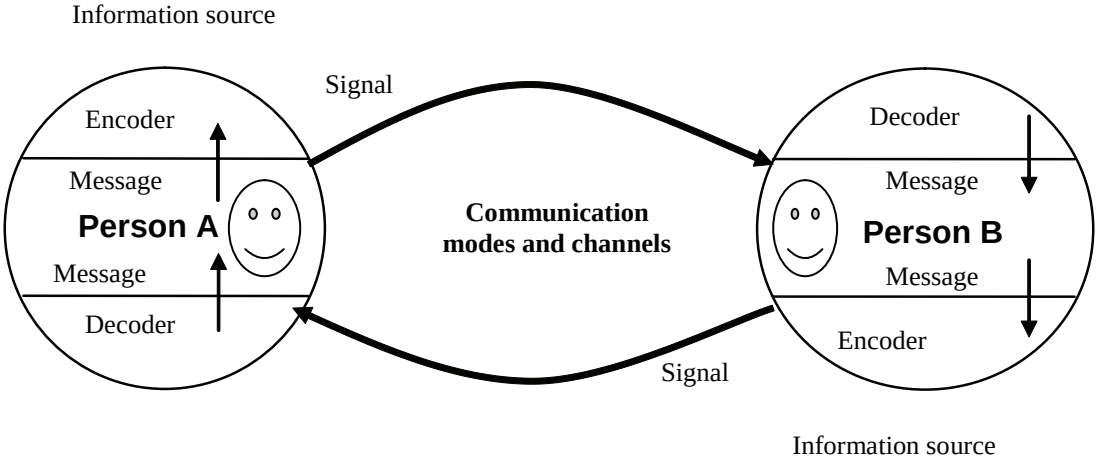
Type 3: In this type, computer-based systems mediate and facilitate the various operations and job roles, such as knowledge-based tasks, repair and maintenance jobs, and so on. Support for this type of application can be called *operation-integrated EPSS.* The emerging technologies involving wearable computers and virtual reality applications supporting repair jobs fall in this class of applications.

Numerous examples of EPSS applications are available in the literature (Banerji, 2003; Dickelman, 2001; Gery, 1991; Hall, 2003). However, detailed discussion of specific EPSS tools is not possible within the confines of this chapter.

INTERACTIVE TECHNOLOGIES FOR COMMUNICATION

Let us now examine multimedia technology as a tool for communication and information transfer. Communication is central to the development of human society and is responsible for all the knowledge that we have accumulated so far. By the word "communication," we mean the process of transmission of data and information from one person to another, which ultimately may lead to knowledge after processing in the mind of the recipient.

Figure 3: Encoder-decoder and channel model of communication



The communication process for information transfer is usually bidirectional. The chain of events starts with a trigger in the mind of the sender (Person A), who, in turn, gives the idea a form by encoding it in a language or expression or picture. The encoded message (signal) is then transmitted to the receiver (Person B), who must have the appropriate decoder to understand the message conveyed by the signal. The receiver may appropriately respond by similarly returning a message to the sender after suitable encoding.

The process of communication, however, continues only if the receiver has the appropriate decoder/encoder. This model is shown in Figure 3. This model can also be easily modified for human–computer communication by replacing "Person B" in the model with a computer. The developments in multimedia and Internet technologies provided the necessary impetus for this evolving human–computer symbiosis utilizing the various communication modes and channels. Their possible applications are limitless, as the technology is under constant evolution.

In the case of human-computer communication, interaction with the computer and the communicative dialogue takes place through some limited modes and channels. The modes are mainly visual, audio, and tactile. Within each mode, there can be various channels. Examples include text, graphics, animation, and video channels in visual mode; voice, sound, and music in audio mode; discrete and continuous tactile interaction modes using keyboard and mouse/joystick, etc. The effectiveness of this communication depends on how well the modes and the media components have been selected and combined. Various interactive technologies are available for this purpose. Detailed discussion on these will be beyond the scope of this chapter. However, the above model is important for conceptualizing and realizing the three types of EPSS discussed earlier.

DESIGN PRINCIPLES

The foregoing discussions on human–task interaction and interactive technologies for communication give us the necessary foundation for appropriate design of interactive multimedia for learning and, particularly, for performance support. Although the complexity of the application domain can vary considerably, we can design an appropriate performance support solution based on a set of 10 fundamental principles and guidelines. These are listed in Table 1. The 10 basic principles formulate the design strategies

Table 1: Basic principles of performance support

Number	Principle	Remarks
1.	**Specify and prioritize** the critical areas of underperformance within the application domain and then **identify appropriate strategies** to improve performance	Suggests a methodology for identification of critical areas of performance deficiency and a top-level approach for their remedy
2.	Attempt to **design mechanization aids and automation tools** to facilitate increases in personal and organizational performance with respect to task-oriented skills	Possible measures for prioritizing tasks could be based on cost, quality, error rate, and task performance time
3.	Identify relevant **generic and application-oriented tools and processes** that will provide on-the-job support and improve task performance	
4.	Attempt to identify and, if possible, eliminate all unnecessary **information blockages and constrictions** within an organization or a work environment	Suggests generic tool sets for information provision, information dissemination, intervention of JIT
5.	Identify an appropriate **combination of media, multimedia, hypermedia, and telecommunications in** order to optimize information flow and interpersonal communications	Training and learning facilities, including eLearning, within a performance support environment
6.	Where a user or employee has an identified skill deficiency, attempt to rectify the situation using **just-in-time (JIT) training and learning techniques**	
7.	Whenever feasible, a performance support system should accommodate **individual learning styles** and thus attempt to maximize its utility for as wide a range of users and task performance situations as is possible	Accommodates the importance of various types of users and their learning styles
8.	Identify appropriate groups of people who have the expertise needed to solve demanding problems and provide the infrastructure necessary to **facilitate group working**	Suggests computer-supported collaborative work (CSCW), the use of intelligent agents, and Knowledge Management
9.	Whenever feasible, attempt to use **intelligent agents** within an EPSS facility in order to **(a) identify the skills needed** for a given task, **(b) locate sources of organizational expertise** relevant to these tasks, and (c) enhance software components	
10.	Attempt to provide facilities to **create a corporate pool of knowledge and skill assets** that can be used to maintain and enhance performance levels	Permeate benefits of performance support right across an organization; create a corporate knowledge pool and skill asset (knowledge capital) that can be made available throughout an organization and is available when needed

for EPSS, including its major supporting components—eLearning and knowledge management (Banerji, 1995).

IMPLEMENTATION APPROACH

Performance means to complete a task such as a piece of work or a duty according to a usual or established method. It also means mastering the task using the most efficient and effective techniques. One aspect of mastering a task using a performance support system is the reliance upon the cognitive partnership between the user and the performance support tool. The important functions and performance measures are as follows:

(1) Reduction in task performance time
(2) Reduction of operational error
(3) Improvement of the quality of task performance
(4) Reduction in cost

These can be achieved through appropriate design of EPSS (Barker & Banerji, 1995; Banerji, 1999a; Gery, 2002).

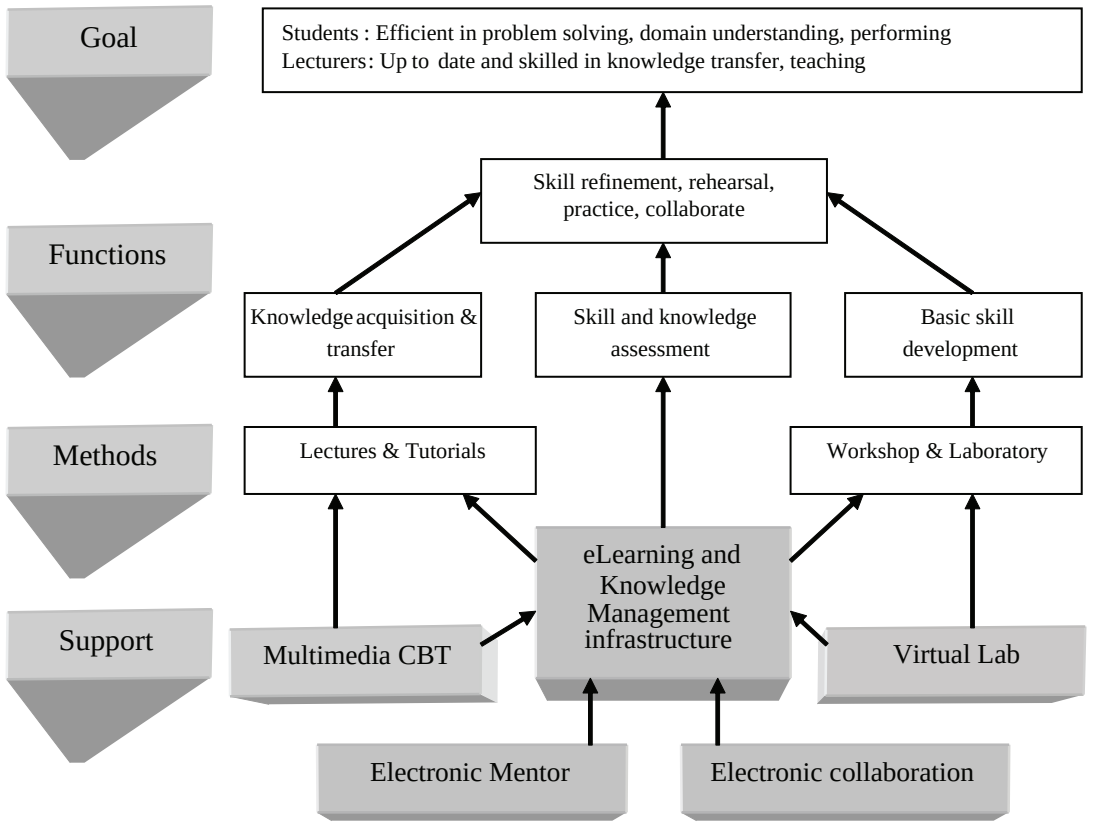
EPSS for Teaching and Learning

The four parameters, time, error, quality, and cost, form the justification for the use of the EPSS approach in any workplace design/redesign. For example, these are equally applicable in any academic institution or corporate university for supporting (a) the students in their learning tasks, (b) the faculty in their tasks of delivering knowledge, and (c) the employees in their management tasks and functions. Let us elaborate one approach.

Despite the advent of powerful, inexpensive, easy-to-use computer technology, the uptake of computer-assisted learning and computer-based training methods within most academic institutions had so far been slow. However, a new wave in the form of *Virtual Classroom*, *Virtual University*, *Web-Based Training* is currently sweeping across most institutions all over the globe. These are clubbed under the term eLearning (or e-learning), which provides opportunities for new modes of information exchange, information transfer, and knowledge acquisition.

It is conceivable that for some time to come, lectures will continue to be the mainstream mechanism for the bulk dissemination of information and knowledge to large groups of students. Given this situation, it is important to address the issue of how best to leverage technology to improve the quality of students' learning experiences and at the same time provide a more effective and efficient framework for the faculty to develop and present

Figure 4: Concept map of support system for teaching and learning



material. One way in which this could be done is to create an electronic performance environment that simultaneously fulfils the needs of both faculty and students. The model of required support for this purpose is shown in Figure 4.

The model shown in Figure 4 is based on the recognition of the currently accepted strategy. It suggests how we can incorporate the new strategy in making the shift (a) from teacher-centered instruction to student-centered learning, (b) from information delivery to information exchange, and (c) from passive learning to active/exploratory/inquiry-based learning.

The distinguishing characteristic of knowledge and skill is that it derives from and finds its meaning in activity. Knowledge is dynamic. Its meaning is constructed and reconstructed as the individual grapples with the use of knowledge through conceptualization, analysis, and manipulation. This naturally has important implications for curriculum development. The objectives of education in any discipline are conventionally attained through (a) classroom training (conceptual understanding), (b) tutorials (analysis), and (c) laboratory practice (practical skill or manipulation).

However, in view of the rapidly changing practices, revitalizing education and training, particularly technical education, has become a matter of concern. This is because of two factors. First, setting up an appropriate up-to-date laboratory is costly and takes time. Second, accessibility of the laboratory is limited to fixed available hours. The existing practices therefore do not support open and flexible learning (Barker, 1996; Baker et al., 1995). Therefore, the central challenge lies in how to provide cost-effective learning opportunities for a larger and more diverse student population. Following on the human activity system model (Figure 1) and the basic principles of performance support system design (Table 1), Figure 4 suggests an approach for a support system for teaching and learning. This can be achieved by incorporating the virtual laboratory, multimedia computer-based training (CBT), including eLearning, knowledge management infrastructure, and the support components made available through the Internet and communication facilities, as shown in Figure 4.

CONCLUSIONS

This chapter described design approaches to assist the knowledge worker of today by leveraging technology to support learning and performance. The basic premise of the approach is incorporating new techniques to deliver just-in-time learning into an EPSS design. Design for this purpose needs sound judgment and decision about pedagogy, which is often the main cause of failure, not the technology. It should be realized that merely hosting Web pages with all the information about the subject is not what eLearning is about. Better learning will not occur if only a conversion of media is effected—from paper to digital.

With sound design, the potentials of interactive multimedia technologies for learning and performing are many. Technology is available now to make learning interesting and activity oriented. It is possible to create low-cost alternatives for learning with active experimentation through virtual laboratories, where learning occurs through practicing and visualizing the concepts. Most importantly, it is possible to make these benefits available in a consistent way to a wider cross-section of people covering a large geographical area.

Gary S. Becker, Nobel laureate and professor of economics and sociology at the University of Chicago, argues the following (Ruttenbur, Spickler, & Lurie, 2000):

The beginning of this century should be called "The Age of Human Capital." This is because the success of individuals and economies succeed will be determined mainly by how effective they are at investing in and commanding the growing stock of knowledge. In the new economy, human capital is the key advantage. (p. 12)

In the knowledge-based economy, organizations as well as individuals need to focus on protecting their biggest asset: their knowledge capital. Therefore, the leaders of companies competing in the knowledge economy have to recognize the importance of efficient knowledge management as well as the importance of developing and enhancing their intellectual capital leveraging technology.

The increasing economic importance of knowledge is blurring the boundaries for work arrangements and the links between education, work, and learning. In this regard, the electronic performance support approach provides a holistic framework for workplace design and redesign.

Of course, the human mind is not going to be replaced by a machine, at least not in the foreseeable future. There is little doubt that teachers cannot be replaced with technology. However, technology can be harnessed as a tool to support the new paradigm. We can derive much gain by adopting information and communication technologies appropriately, especially as we look for new solutions to provide the knowledge worker with immediate learning opportunities.

REFERENCES

Banerji, A. (1995). Designing electronic performance support systems. *Proceedings of the International Conference on Computers in Education (ICCE95)* (pp. 54–60). Singapore, December 5–8.

Banerji, A. (1999a). Performance support in perspective. *Performance Improvement Quarterly, 38*(7). Retrieved from the World Wide Web: http://www.pcd-innovations.com/piaug99/PSin-Perspective.pdf

Banerji, A. (1999b). Multimedia and performance support initiatives in Singapore Polytechnic. *SP Journal of Teaching Practices.* Retrieved from the World Wide Web: http://www.vc.sp.edu.sg/journals/journals_intro.htm

Banerji, A. (Ed.) (2001). The world of electronic support systems. Retrieved February 6, 2004 from the World Wide Web: http://www.epssworld.com/

Bare, J., & Meek, A. (1998). *Internet access in public schools* (NCES 98-031). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Barker, P., & Banerji, A. (1995). Designing electronic performance support systems. *Innovations in Education and Training International, 32*(1), 4–12.

Barker, P., Banerji, A., Richards, S., & Tan, C. M. (1995). A global performance support for students and staff. *Innovations in Education and Training International, 32*(1), 35–44.

Boyer, R. S., Feijen, W., et al. (2002). In memoriam Edsger W. Dijkstra 1930–2002. *Communications of the ACM, 45*(10), 21–22.

Clark, R. E. (Winter 1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920.* New York: Teachers College Press.

Dede, C. (1998). *Six challenges for educational technology.* Retrieved from the World Wide Web: http://www.virtual.gmu.edu/SS_research/cdpapers/ascdpdf.htm

Dickelman, G., & Banerji, A. (1999). Performance support for the next millennium: A model for rapidly changing technologies in a global economy. HCI 99 Conference, Munich.

Dickelman, G. J. (Ed.) (2003). EPSS design contest awards. Retrieved February 6, 2004 from the World Wide Web: http://www.pcd-innovations.com/

Gery, G. (2002). *Performance support—Driving change* (pp. 24–37). The ASTD E-Learning Handbook, Ed. Allison Rossett. New York: McGraw Hill.

Gery, G. J. (1991). *Electronic performance support systems: How and why to remake the workplace through the strategic application of technology.* Boston, MA: Weingarten Publications.

Hall, B. (Ed.) (2003). Retrieved February 6, 2004 from the World Wide Web: http://www.brandonhall.com

Licklider, J. C. R. (1960). Man–computer symbiosis. *IRE Transaction of Human Factors in Electronics*, *HFE*-1(1), 4–11.

Marino, T. (2001, July/August). Lessons learned: Do you have to bleed at the cutting edge? *The Technology Source*. Retrieved from the World Wide Web: http://ts.mivu.org/default.asp?show=article&id=860#options

McArthur, K. E. (2000). *Teachers use of computers and the Internet in public schools* (NCES 2000090). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Reeves, C. T. (1998). *The impact of media and Technology in Schools.* A research report prepared for the Bertelsmann Foundation. Retrieved from the World Wide Web: http://www.athensacademy.org/instruct/media_tech/reeves0.html

Ruttenbur, B. W., Spickler, C. G., & Lurie, S. (2000). *eLearning the engine of the knowledge economy.* Retrieved from the World Wide Web: www.morgankeegan.com; http://www.masie.com/masie/researchreports/elearning0700nate2.pdf

Tannenbaum, R. S. (1998). *Theoretical foundations of multimedia* (Chapter 5). New York: W.H. Freeman & Co. Computer Science Press.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, pp. 47-59, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.7
Planning for Multimedia Learning

Patrick J. Fahy
Athabasca University, Canada

ABSTRACT

Multimedia tools, applied with awareness of the realities of organizational culture, structures and finances, have been shown to enhance the performance of learning systems. If some predictable pitfalls are avoided, and proven pedagogical design principles and appropriate vehicles (including the Internet) are used effectively, multimedia can permit greater individualization, in turn fostering improved learning, learner satisfaction, and completion rates.

INTRODUCTION

Effective uses of multimedia in open and distance learning (ODL) depend upon various factors, some intrinsic to the media themselves, and others related to the differing pedagogic tasks and organizational environments into which these tools are introduced. For those planning use of multimedia, it may be valuable to consider the likely impacts of these tools on teaching and learning practices and outcomes, and on organizational structures and processes, as they are likely to be different in scope and magnitude from those of traditional instructional innovations.

This chapter discusses some of the characteristics of multimedia in relation to basic pedagogic tasks and organizational realities. The goal is to alert new users to issues that often arise in multimedia implementations and to assist experienced users in assessing their strategies, by outlining some fundamental considerations commonly affecting implementation of multimedia. Both new and experienced technology users will hopefully find the discussion useful for reflecting on options, and anticipating potential pedagogic and administrative challenges, as they move from simpler to more complex combinations of media for teaching.

The chapter begins with a discussion of the term *multimedia*, including a review of some of the characteristics (including common pedagogic benefits and potential issues) of specific media. Based on this analysis, some of the conditions under which multimedia might readily support

learning tasks are explored. Finally, the impact of multimedia as an innovation on aspects of organizational culture (including structure and finances) are addressed.

Defining Multimedia

While the term "multimedia" has not always been associated with computers (Roblyer & Schwier, 2003, p. 157), there is no doubt that it is the merging of increasingly powerful computer-based authoring tools with Internet connectivity that is responsible for the growing interest in and use of multimedia instruction, in both distance and face-to-face environments. This trend is encouraged by growing evidence that well-designed online delivery, regardless of the media used, can improve retention, expand the scope and resources available in learning situations, and increase the motivation of users (Fischer, 1997; Bruce & Levin, 1997; Mayer, 2001). For these reasons, the term "multimedia" is now firmly associated with computer-based delivery, usually over the Internet and accompanied and supported by interaction provided via some form of computer-mediated communication (CMC).

Definitions of multimedia vary in particulars but tend to agree in substance. Mayer (2001, p. 1) defined multimedia learning simply as "presentation of material using both words and pictures." Roblyer and Schwier (2003) observed that definition is problematic, because it is increasingly difficult to distinguish multimedia from other tools with which it seems to be converging. They also note that multimedia have sometimes been defined simplistically by the storage devices they employ, e.g., CD-ROM, videodisc, DVD, etc., a practice they regard as clearly inadequate. Roblyer and Schwier offered this definition of multimedia: "A computer system or computer system product that incorporates text, sound, pictures/graphics, and/or audio" (p. 329). They added that the multimedia implies the purpose of "communicating information" (p. 157).

In keeping with the above, in this chapter, the term "multimedia" refers to the provision of various audio and video elements in teaching and training materials. Usually, the delivery of the media is by computer, and increasingly, it involves the Internet in some way, but the storage and delivery devices, as noted above, are secondary to the forms of the stimuli that reach the user. The definition assumes that media are used, but it does not address such design issues as choice of specific media for differing pedagogic purposes and levels of user control.

Basic to considering how specific media contribute to the effectiveness or ineffectiveness of multimedia is a brief discussion of the available research on technology in learning. Multimedia technologies invariably consist of media with effects on learning that have been studied before, making this knowledge pertinent and applicable here (Saettler, 1990).

MEDIA AND LEARNING

Specific Media Characteristics

For some time, media have been used with more traditional delivery methods (lectures, tutorials) to support essential teaching objectives, such as the following (Wright, 1998):

- Clarifying and illustrating complex subjects
- Adapting to individual learning styles
- Improving retention and aiding recall
- Reaching nonverbal learners

Debates have occurred over the precise role of media in learning. The fundamental disagreement between Clark (1983, 1994) and Kozma (1994) about media and learning is familiar historically and need not be repeated here. It seems clear that Mayer's (2001) views of multimedia (discussed later) clearly support one point made in that debate, that of the "interdependence" of presentation

media and delivery methods in certain circumstances, especially in collaborative situations, and where higher-order learning is an objective (Crooks & Kirkwood, 1988; Juler, 1990; Koumi, 1994). As Berge (1995, p. 23) concluded, and as has been documented by Mayer (2001), "Some media channels promote particular interactions, and other channels can hinder that same type of interaction."

While the potential for successful high-level learning outcomes is present in media use, a persistent problem in multimedia applications has been failure to achieve more than low-level learning outcomes (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). Helm and McClements (1996) commented critically, "Interactivity in the context of multimedia often refers to the learners' ability to follow hypertext links or stop and start video clips…. Much of what passes for interactivity should really be called feedback" (p. 135). These are serious criticisms, justifying Mayer's (2001) advice, "Instead of asking which medium makes the best deliveries, we might ask which instructional techniques help guide the learner's cognitive processing of the presented material" (p. 71).

The varying characteristics of different presentation media and modes, and their implications for learning, have direct implications for the design of multimedia strategies and materials. *Sound* can supplement visual information and can be used to attract attention, arouse and hold interest, provide cues and feedback, aid memory, and provide some types of subject matter (heart or machinery sounds, voice clips). *Music* can be used to augment feedback, grab attention or alert users, and support the mood of a presentation. *Synthetic speech*, while useful for handicapped users, is less effective if too mechanical sounding. Szabo (1998) concluded that achievement gains due to audio are "weak or non-existent." He added that where benefits are seen, they tend to accrue to the more highly verbal learners. Problems with development costs and bandwidth for delivery of audio can also be significant (Wright, 1998; Szabo, 1998).

Graphics and color can be used for various purposes, from simple decoration to higher-level interpretation and transformation (helping the observer to form valid mental images) (Levin, Anglin, & Carney, 1987). Research has shown that realism and detail are not critical in graphics and may, in fact, extend learning time for some users; relevance is more important than detail (Szabo, 1998). Color may also distract some learners, unless it is highly relevant to instruction. A significant proportion of individuals (especially men) have some degree of color-blindness, suggesting that color should be placed under the control of the user where possible. The best contrasts are achieved with blue, black, or red on white or white, yellow, or green on black.

Animation can sometimes shorten learning times by illustrating changes in the operation or state of things; showing dangerous, rapid, or rare events; or explaining abstract concepts. For some, animation increases interest and holds attention better than text or audio, and the resulting learning seems to be retained (Szabo, 1998). Overall, however, research indicates that well-designed and imaginative verbal presentations may be capable of producing similar outcomes (Rieber & Boyce, 1990), leading to the conclusion that animation may not possess many unique instructional capabilities.

Video (motion or sequences of still graphics) can be used to show action and processes and to illustrate events that users cannot see directly or clearly in real time. Video, when used skillfully and artistically, can also emotionally move observers and can produce impacts affecting attitudes similar to in-person observation of real events.

Hypermedia is the linking of multimedia documents, while *hypertext* is the linking of words or phrases to other words or phrases in the same or another document (Maier, Barnett, Warren, & Brunner, 1996, p. 85). Hypertext and hypermedia may be difficult to distinguish and

increasingly difficult to separate from other applications of multimedia (Roblyer & Schwier, 2003). When paired with plain text, hypertext has been shown to be a cost-effective way to extend text's information-conveying capabilities, especially for more capable learners. Szabo (1998) suggested that hypertext should be used more to provide access to information than for actual teaching, in recognition of the need for hypertext materials to be placed in context for maximum impact (especially for less experienced or less capable learners).

Hypermedia is a particularly promising form of multimedia materials designed for ODL (Maier, Barnett, Warren, & Brunner, 1996, p. 85; Roblyer & Schwier, 2003). With advances in hardware, software, and human–computer interfaces, it is now technically feasible to use hypermedia systems routinely in online teaching. Dozens of hypertext and hypermedia systems exist, with most offering three basic advantages:

- Huge amounts of information from various media can be stored in a compact, conveniently accessible form, and can easily be included in learning materials.
- Hypermedia potentially permit more learner control (users can choose whether or when to follow the available links).
- Hypermedia can provide teachers and learners with new ways of interacting, rewarding learners who developed independent study skills and permitting teachers to be creative in how they interact with learners (Marchionini, 1988, p. 3).

There are potential problems, too, in learning with hypermedia, related to the volume and structure of all information found on the Web. The vast amounts of information available can overwhelm the learner, especially if structure is inadequate or procedures such as searches are not skillfully refined, allowing learners to "wander off" and become engrossed in appealing but irrelevant side topics. Learners who do not have independent study skills may not be able to manage the complexity of hypermedia. This problem may not be immediately evident, however, because they *appear* to be engaged and on task, sometimes deeply so.

Other potential problems in teaching with hypermedia include some unique to this medium and others common to all learning situations that require specific skills or make assumptions about learner attributes and characteristics:

- Hypermedia require basic literacy skills. While this may change as increasing bandwidth makes audio and video available, presently, the Internet and its multimedia products rely heavily on text.
- A related problem is that interacting with hypermedia and multimedia requires keyboard and mouse skills, as well as understanding and manipulating function keys. The computer illiterate, the unskilled, or the physically handicapped may be affected.
- More broadly, accessing hypermedia and multimedia requires computer use, including sitting in front of the machine and making sense of its cues and displays. Those with vision, concentration, coordination, or mobility problems, or those distracted or confused by the intense stimulation of colors, animation, sound, etc., may be penalized.

The above specific features of media have been shown to affect their usefulness for teaching and learning. In addition to the limitations of media, a key point here is the importance of historical media research to the present discussion: multimedia are *media*, and the view taken in this chapter is that knowledge previously gained about their impact on learning is still highly applicable.

Media Characteristics, Teaching Conditions, and Learning Outcomes

When media are used together, their effects can interact, sometimes unpredictably. With media, "more is not necessarily better."

There is as yet little thorough research on multimedia technologies to inform design and implementation decisions; use of previous research may help guide present practice. What follows is a discussion of some key didactic purposes to which media may apply, followed by some remarks about the Internet as a base for multimedia delivery.

Evaluations have shown that a fundamental benefit to students from the best uses of technology in teaching is a more systematic approach to the individualization and customization of instruction (Massy & Zemsky, 1999). Properly designed, a technology-based learning environment provides students with more options than are typically available in traditional learning situations, in content, pace, preparation, and review of prerequisites, and for activities such as collaboration, consultation, and testing/evaluation. These are objectives that have long been recognized as pedagogically essential (Zimmerman, 1972; Mezirow & Irish, 1974; Kemp, 1977; Dede, 1996; Roblyer, Edwards, & Havriluk, 1997). Among the benefits of technology delivery are the potential for less required training time; greater mastery and better transfer of skills; more consistency in delivery of content (a particularly important outcome of skill training); and greater student persistence, completion, satisfaction, collaboration, and self-direction (Grow, 1991; Moore, 1993). In some situations, experience has shown that highly self-directed students may be able to undertake and complete advanced studies with little or no direct assistance or intervention from the institution, increasing efficiency through the "unbundling" of learning from direct teaching (Massy & Zemsky, 1999, pp. 2–3). In the best examples, technologies increase learning, enhance learner satisfaction, stabilize costs, and raise the visibility and appeal of (and potential revenues from) existing programs (Oberlin, 1996).

While positive effects are possible in teaching with media, they are not automatic. Internal consistency of objectives is critical: multimedia technologies must be congruent with the organization's learning model and actual teaching practices, as well as with students' expectations and capabilities for autonomy and self-direction (Grow, 1991). If tools are chosen for their technological capabilities alone, there is a risk of failing to fit with the organizational environment (Helm & McClements, 1996; Mayer, 2001; Welsch, 2002), resulting in potentially disastrous technology implementation "mistakes" (Quinn & Baily, 1994).

Despite differing characteristics, useful online training technologies have in common the effect of bringing the student into timely and productive contact with the tutor, the content, and peers, thereby reducing the "transactional distance" in distance learning, the communications gap or psychological distance between geographically separated participants (Moore, 1989; Chen & Willits, 1998). The differences in how various media accomplish their effects are important to their potential usefulness. Figure 1, for example, compares instruction delivered by human and technological means (Fischer, 1997).

Illustrated in Figure 1 are some of the trade-offs inherent in the decision to use teaching media, as opposed to traditional forms of delivery alone. If a critical value for a program is met by tutor-based delivery, and resources are plentiful, it may be chosen without regard for cost. Where economy is important, however, the "best" delivery solution may not be affordable; a less costly but still adequate solution may have to be chosen. (This was the purpose of Bloom's [1984] "two-sigma" challenge, to find a teaching medium as effective as one-on-one tutoring. The search, of course, continues with multimedia.) Analysis such as the above may assist in identifying the trade-offs involved in the choice of one medium or technology over another and may suggest compensating strategies to improve the effectiveness of whatever tool is chosen (Wolfe, 1990).

Besides cost and accessibility (Bates, 1995), another issue in selecting media is the type of

Figure 1: Comparison of characteristics of human- and technology-based instruction

Training element	Human-delivered training	Technology-based training
Planning and preparation	Able to design training to correspond to the training plan; able to monitor consistency	Must be systematically designed to conform to the training plan
Expertise	Presenters hired from industry usually represent the most current knowledge and highest expertise	Must be designed to conform to industry standards; currency with standards must be maintained
Interactivity	Instructors tend to train the group, ignoring individual needs	Able to focus on individual needs in content, pacing, review, remediation, etc.
Learning retention	Retention rates vary	Can be up to 50% higher than instructor-led group training
Consistency	Instructors tend to adapt to the audience, sacrificing consistency	Rigorously maintains standards but may also be designed to adapt to learner's performance or preferences
Feedback, performance tracking	Human instructors especially good at constant, ongoing evaluation, response to trainee performance	Better at keeping records and generating reports, but designing cybernetic systems to adapt instruction based on feedback is costly, complex

experience or learning outcomes intended by the training (DeSanctis & Gallupe, 1987). Picard (1999), for instance, sees the key contribution of media as their ability to promote *relationship building*, and not merely *information exchange*, in work or learning.

From Figure 2, we see the following:

- When relationship-building and information exchange needs are both low, audio media alone may suffice.
- When both relationship-building and information-exchange needs are high, audio, video, and information exchange (including text) should all be present.
- Relationship-building is enhanced by combining audioconferencing and video together with data, especially text. (Text alone has substantial relationship-building capabilities, as anyone who has ever had a pen pal, or exchanged love letters, can attest.)

In relation to learning, technologies have potential directly to address common teaching tasks. In Figure 3, the views of several theoreticians regarding tasks or conditions essential to learning are compared. Two points should be noted in this comparison: (a) there is considerable apparent agreement among authorities on elements essential to effective teaching and learning, and (b) there appear to be obvious roles for multimedia in supporting some of these tasks.

A broader point in this discussion is made in Figure 3: technologies have capabilities to assist in specific teaching tasks, if used within their identified limitations as presentation and delivery media. The purpose of research on media is to identify characteristics (capabilities and limitations) that can then be applied in the ID phase, thus avoiding use of the wrong tool for a specific pedagogical purpose. Previous media research can be useful in identifying multimedia implementations able to supply or support the following:

- **Instruction—**CAL (computer-assisted learning), including various types of simulations, can be used, supported by varieties of CMC (e-mail, synchronous and asynchronous IP-audio- and IP-videoconferences, text-chat, file exchanges, and data access).

Figure 2: Relation of data, audio, and video technologies to information exchange and relationship-building outcomes

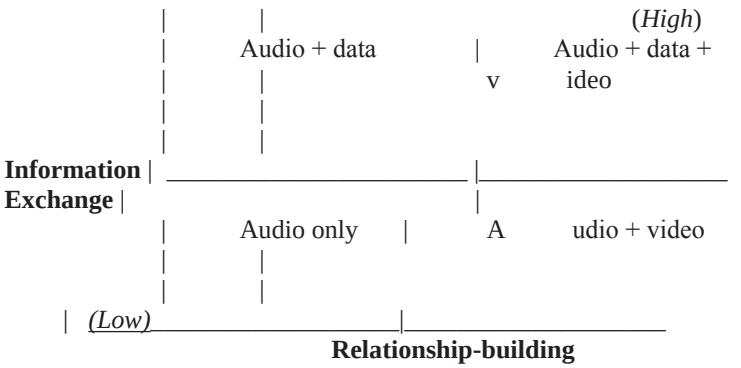


Figure 3: Comparison of models of effective teaching and learning: roles for multimedia

Bloom (1984)	**Chickering & Gamson (1989)**	**Gagne (1985)**	**Joyce & Weil (1980)**	**Moore (in Garrison, 1989)**
Tutorial instruction	Student–faculty interaction	Presenting new material; describing objectives; gaining learner's attention	Presenting stimuli, objectives; sequencing learning tasks	Communicating in an understandable manner
Reinforcement	Student–faculty interaction	Recalling previous learning; enhancing retention and recall	Increasing attention; promoting recall	General support
Corrective feedback	Proper feedback	Providing feedback on performance	Prompting and guiding	Feedback
Cues and explanations	Student–faculty interaction	Learning guidance	Prompting and guiding	Guidance
Student participation	Active learning; student reciprocity and cooperation	Student performance	Evoking performance	Active involvement
Time on task	Time on task; communicating high expectations	Assessing performance		
Assessing and enhancing learner's reading and study skills	Respecting diverse ways of learning			

- **Reinforcement, corrective feedback, and cues and explanations—**CAL and, especially, CML (computer-manager learning) can be useful.
- **Participation, engagement, time-on-task—S**trategies for collaboration and cooperation with peers and authorities include various forms of problem-based learning, using Internet-based communications tools. Motivational advantages are gained from the scope of access and the immediacy of interaction provided by the Web.
- **Assessing and respecting diverse learning styles, preferences—**Though not cited by all the authorities in Figure 3, this may be one of the most powerful arguments for multimedia delivery. (As Fletcher [1992] recognized more than a decade ago, individualization is both "a morale imperative and an economic impossibility"—unless, it is argued here, use is made of well-designed multimedia resources.)

As noted earlier, technologies vary in their immediacy and interpersonal impact. For example, video affects the likelihood and, according to some research, the speed with which relationships will grow in mediated interaction, while simple data exchange may do little to promote relationships in virtual work teams (Walther, 1996; Picard, 1999). The objectives of the instruction should dictate the media to be used and must be grounded in the media's demonstrated capabilities; the choice of media thus both affects and reflects the relative emphasis on different desired learning outcomes.

Multimedia and the Internet

Multimedia are increasingly associated with the Internet, which offers both delivery advantages and challenges to users: advantages arise from the Internet's enormous capacity to link and interconnect, but there are potentially serious problems related to lack of inherent structure and tutor control (Thaler, 1999; Stafford, 1999; Campbell, 1999). Advantages of the Internet for teaching, under ideal conditions, include the following (Heinich, Molenda, Russell, & Smaldino, 1996, p. 263):

- ***Engrossing:*** The opportunity for deep involvement, capturing and holding learner interest.
- ***Multisensory:*** The incorporation of sounds and images along with text (but see Mayer's [2001] *multimedia principles,* below, regarding the limits of sensory channels).
- ***Connections:*** Learners can connect ideas from different media sources, for example, connecting the sound of a musical instrument with its illustration.
- ***Individualized:*** Web structure allows users to navigate through the information according to their interests and to build their own unique mental structures based on exploration.
- ***Collaborative creation:*** Software allows teachers and learners to create their own hypermedia materials; project-based learning provides opportunities for authentic collaboration.

Some of the more common problems with the Internet for teaching, and as a platform for multimedia delivery, are as follows (Heinich et al., 1996, p. 263):

- ***Getting lost:*** Users can get confused, or "lost in cyberspace."
- ***Lack of structure:*** Those whose learning style requires more structure and guidance may become frustrated. Some less-experienced or less well-disciplined users may also make poor decisions about how much information they need.
- ***Noninteractive:*** Programs may simply be one-way presentations of information with no specific opportunities for interaction or practice with feedback. A further problem is that, due to poor design, what may be intended as *interaction* is sometimes more accurately called *feedback* (Helm & McClements, 1996).
- ***Time-consuming:*** Because they are nonlinear and invite exploration, hypermedia programs tend to require more time for learners to reach prespecified objectives. Because they are more complex than conventional instructional materials, hypermedia systems require more time to master ("Workers find," 2000).
- ***Bandwidth***: This continues to be a critical barrier to Web-based multimedia use for some potential users. While broadband availability is increasing worldwide (*PC Magazine*, 2003), especially outside North America ("Where's the broadband boom?," 2002), online speeds still prevent many users from accessing multimedia efficiently or reliably (Howard, 2001; Miller, 2002).

The above inherent limitations of the Internet as a multimedia delivery tool arise from its very nature. In order for these limitations to change, the Internet would have to become more structured, limiting user choices. This is unlikely, as these changes would make the Web a very different entity from what it is today (Greenaway, 2002).

Planning Issues with Multimedia

Design and Development Principles

The potentials and challenges discussed above underscore the importance of planning and design in the implementation of multimedia. Fortunately, research offers principles that can guide instructional designers and instructors in the development and use of multimedia. Mayer's (2001) work is particularly useful. His examination of the impact of multimedia on learning, based on how the human mind works to process verbal and visual information (p. 4), has produced important insights about media and learning, including the following:

- Words and pictures, although qualitatively different, complement one another and promote learning, *if* learners are successful in mentally integrating visual and verbal representations (p. 5).
- True learning is more a process of knowledge construction than information acquisition (p. 12).
- Deep learning is evidenced by retention *and* transfer (lack of which indicates no learning, or merely superficial rote learning) (pp. 5, 16–17).

In Mayer's model, there are three key assumptions underpinning a cognitive theory of multimedia learning: (a) humans have dual channels for processing input as part of learning, the *visual* and the *auditory*; (b) while the two channels exist in most people, humans are limited in the amount of the information they can process in each channel at one time; and (c) learners must actively process information and experience as part of learning, by a process that includes attending to relevant incoming information, organizing selected information into coherent mental representations and integrating mental representations with other knowledge (p. 44).

Mayer (2001, p. 41) concluded that successful learning requires students to perform five actions, with direct implications for the design of effective multimedia instruction:

1. Select relevant words from the presented text or narration.
2. Select relevant images from the presented illustrations.
3. Organize the selected words into a coherent verbal representation.
4. Organize selected images into a coherent visual representation.
5. Integrate the visual and verbal representations with prior knowledge.

Mayer articulated seven principles useful for guiding the design of multimedia instruction. Under these principles, students have been shown to achieve greater retention and transfer (Mayer, 2001, p. 172):

1. **Multimedia principle**: Students learn better from words and pictures than from words alone.
2. **Spatial contiguity principle**: Students learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.
3. **Temporal contiguity principle**: Students learn better when corresponding words and pictures are presented simultaneously rather than successively.
4. **Coherence principle**: Students learn better when extraneous words, pictures, and sounds are excluded rather than included. ("Extraneous" can refer either to *topical* or

conceptual relevance, with the latter being more important.)

5. **Modality principle**: Students learn better from animation and narration than from animation and on-screen text. (This principle assumes use of a *concise narrated animation*, text that omits unneeded words.) (See p. 135.)
6. **Redundancy principle**: Students learn better from animation and narration than from animation, narration, and on-screen text. (This principle is based on *capacity-limitation hypothesis*, which holds that learners have limited capacity to process material visually and auditorily [p. 152]. Eliminating redundant material results in better learning performance than including it [p. 153]).
7. **Individual differences principle**: A particularly important finding is that design effects are stronger for low-knowledge learners than for high-knowledge learners, and for high-spatial learners than for low-spatial learners (p. 184).

The above are examples of design principles under which learning *may* be enhanced by the use of various display or delivery media. Principles such as these are particularly important, as they are research-based and tested (Mayer, 2001). Any design principles adopted should meet similarly stringent empirical tests.

Multimedia, Productivity and Performance

The previous discussion suggests that multimedia implementation, while potentially valuable to learning, requires strategic planning to exploit pedagogic possibilities and avoid the pitfalls of misapplication. The point has further been stressed that the existing literature on technology-based learning is applicable to multimedia planning, especially the known pedagogic and representational characteristics of individual media identified in actual learning situations. There are nonpedagogic considerations, too, related to organizational impacts and various costs from the use of multimedia.

A realistic decision to incorporate multimedia in ODL should recognize that multimedia, like most technologies, are unlikely initially, or perhaps ever, to save the organization time or money (Quinn & Bailey, 1994; Burge, 2000; Cassidy, 2000). In fact, multimedia may in the short-term increase operational complexity, create "organizational chaos" (Murgatroyd, 1992), and promote time-wasting behaviors by users throughout the organization (Laudon, Traver, & Laudon, 1996; Fernandez, 1997; Evans, 1998; Fahy, 2000; Dalal, 2001). The early effects of multimedia, like other technologies in complex organizations, may well include *lower* organizational productivity (Black & Lynch, 1996).

Another caveat is financial: the economics of technologies generally suggest that the total cost of ownership (TCO) of multimedia technologies will constantly rise (Oberlin, 1996), and that no genuine cost savings may ever actually be achieved by some users (Welsch, 2002). The rationale for adopting multimedia technologies, therefore, is more related to *performance* enhancements, such as greater flexibility, improved learning, and higher satisfaction and completion rates for users, than to cost savings (Oberlin, 1996; Daniel, 1996; Fahy, 1998).

This point is significant, because, historically, technology users have sometimes confused performance and productivity outcomes in technology implementations, underestimating the costs and long-term impacts of technology, while, to the detriment of realistic expectations, overestimating and overselling possible productivity benefits (Dietrich & Johnson, 1967; McIsaac, 1979; Mehlinger, 1996; Strauss, 1997; Lohr, 1997; Wysocki, 1998; Greenaway, 2002; Hartnett, 2002). For the future of multimedia, avoiding these kinds of mistakes is critical: unrealistic expectations produce disappointment, and may result in skepti-

cism among instructors and managers about the value of educational innovation generally, and educational technologies in particular ("Nothing travels through an educational vacuum like a technological bandwagon.")

Organizational Issues in Multimedia Adoption

Realistic expectations of multimedia require compatibility with the adopting organization's culture, structure, and finances (Welsch, 2002).

The culture of any organization includes its various values, beliefs, myths, traditions, and norms, as well as its historic practices and typical ways of doing business (including how it adopts or rejects innovations). Organizational culture may present the most serious challenges to those responsible for the strategic planning (Rogers, 1983; Stringer & Uchenick, 1986), including the problem of distinguishing whether any resistance encountered is due to simple unwillingness or to real inability (Welsch, 2002). (In the latter case, resistance may be rationale and appropriate, a sign that conditions are not right for an innovation to succeed.) The problem is thought to be particularly acute in slow-to-change enterprises such as public education (Senge, 1990).

Another problem for adoption of complex innovations such as multimedia is the attitude in some organizations that training is an optional activity (Gordon, 1997). Ironically, it is technologically illiterate managers and administrators who most often resist training initiatives, both for themselves and their staff, to avoid embarrassment in an area in which they know their expertise is not as great as their subordinates'. The needs analysis stage of planning is the best place to assure that cultural issues like these are recognized and evaluated in advance.

Planning for multimedia implementation need not be timid. The needs assessment should carefully distinguish *climate* from culture and respond accordingly. Climate consists of the commonly held viewpoints and opinions in the organization, directly influenced by widely recognized measures of organizational health and success, such as enrollment or student achievement and performance relative to competitors. Climate is more "constructed" and temporary than culture, based upon elements such as student and staff perceptions of how well the organization is performing its fundamental tasks. By its nature, climate is more manageable than culture. Managers, by their reactions to external developments, can influence how staff members interpret the outside events that may shape climate. Climate is an area in which planning can have an impact, through the efforts of planners to influence the internal recognition and interpretation of outside events.

In addition to culture, structural factors within the organization may also affect multimedia innovations. The presence and adequacy of the required technological infrastructure, including software, hardware, communications, and networking systems, should be assessed. Personnel in the form of knowledgeable management, maintenance, training and support staff, key consultants, and cost-effective contract help are also critical structural elements. If not already provided for, the costs of system upgrades and ongoing maintenance (including initial and recurrent training for staff) should be assessed in a structural review, preceding the introduction of multimedia systems. Ongoing costs should be identified in budget projections.

Finances are a vital part of any multimedia adoption in ODL. Assessing organizational finances also introduces complexity into the planning process, as costs are inherently difficult to predict accurately, sometimes even to identify completely. While precise accuracy in cost analysis may be difficult, potential purchasers of technologies should be aware that, as noted above, the total cost of ownership of multimedia technology will likely be well above the purchase price, exceeding the purchase price by many times (Oberlin, 1996; Black & Lynch, 1996;

Khan & Hirata, 2001; Welsch, 2002). Using a definition of productivity as the ratio of benefits to costs (Massy & Zemsky, 1999), the high cost of a technology may not be disqualifying if the payback is clear. Costs alone do not necessarily change the justification for a technology, but they could constitute a shock to an organization that has not adequately anticipated them.

Part of the rationale for investing in multimedia is the fact that technology provides flexibility: technologies are more scalable than human resources, *if* this aspect is exploited in the organizational vision. In general, scalability means that program growth may be more easily accommodated with technology than without it; costs do not escalate in line with growth as they do where enrollment increases are borne strictly by hiring more instructors and support staff. Multimedia resources may be augmented or trimmed without reference to collective agreements or other commitments. Another difference is that technologies such as multimedia tend to become more efficient the more use is made of them, lowering the break-even point and increasing their efficiency (Matkin, 1997; Harvard Computing Group, 1998; Watkins & Callahan, 1998).

The decision to acquire technology is fundamentally a strategic one, because technologies are means to various ends. Bates (1995) suggested that *accessibility* and *cost* are the two most important discriminators among technologies, and thus the most critical criteria in a technology acquisition process. A decision to "build" or develop a new multimedia technology option should bear in mind that there is now a rapidly increasing amount of available software (Gale, 2001). A careful analysis of needs and a search of available options should be performed, especially before a decision to develop is authorized, as even professional programming projects, in general, often end in failure (Girard, 2003), and instructors who lack special instructional design (ID) training are particularly apt to become bogged down in the development of ultimately mediocre materials (Grabe & Grabe, 1996).

Another factor in assessing the financial viability of various multimedia technologies is the potential target audience in relation to the expected costs of production. Perry (2000) cautioned that custom multimedia training courseware will likely not be cost-effective for fewer than 1,000 users. Bates (1995, 2000) also offered figures and usage considerations that help with the assessment of costs and benefits. Costs and time frames can be formidable: Szabo (1998) reported nearly a fourfold range (40 to 150 hours per hour of instruction) for development of very basic computer-assisted learning (CAL) in health education, and another study reported that a 6 hour module in weather forecasting, involving a production team of instructional designers, meteorologists, hydrologists, graphics artists, media specialists, computer scientists, and SMEs, consumed a year and cost $250,000 (Johnson, 2000).

CONCLUSIONS

Presented in this chapter was a discussion of factors (inherent, pedagogic, and organizational) that may impact planning for multimedia use. The suggestion here is that multimedia are more likely to affect pedagogical performance (how well the program or the organization does its work) than productivity (measured by profitability). In planning for multimedia implementation, it was suggested, performance outcomes should be the focus (improvements in quality of service, as measured by timeliness, accessibility, convenience, and responsiveness of program offerings and supports), rather than "bottom-line" outcomes.

Strategic planning in the form of ID promotes proper uses of multimedia technologies, especially (at the awareness and adoption stages). The best pedagogical arguments for use of multimedia technologies (providing more learner convenience, satisfaction and success) may be compelling enough, but problems in relation to existing organizational culture, structure and finances should

not be overlooked. The adoption process includes distinguishing climate factors from culture (the former being more amenable to influence by effective leaders); considering the needs of affected groups in planning; acknowledging and respecting users' expectations; providing existing managers with training, so they can provide effective leadership; accurately assessing existing and needed technical resources; avoiding overselling potential benefits, thus keeping expectations realistic; and selecting, adapting, or (rarely) building products on the basis of demonstrable advantages, especially accessibility and costs.

Pedagogically, the principal contributions of multimedia technologies in teaching and training are likely to be increased flexibility, resulting in greater learner access and convenience, and more choices to users, including self-pacing, individualization, customization, and learner control. Positive impacts such as these on aspects of the teaching process can be anticipated, but problems should also be expected; media selection usually involves trade-offs, and the losses and gains in the choice of one delivery or presentation medium over another should be acknowledged.

For instructional designers, principles exist to guide development of multimedia. Among the most useful of these are the multimedia principles that address design issues such as contiguity, redundancy, coherence, and choices of delivery modes (Mayer, 2001). Adoption of these principles would, in general, likely result in "lean" multimedia design, with use of audio-textual and visual-pictorial elements based more directly upon empirical evidence about how these actually impact learning, rather than upon their technical features alone. Though perhaps less technologically elegant, such implementations promise to be more pedagogically effective and organizationally compatible.

REFERENCES

Bates, A. W. (1995). *Technology, open learning and distance education*. New York: Routledge.

Bates, A. W. (2000). *Managing technological change*. San Francisco: Jossey-Bass Publishers.

Berge, Z. (1995). Facilitating computer conferencing: Recommendations from the field. *Educational Technology*, January–February, pp. 22–30.

Black, S., & Lynch, L. (1996). Human-capital investments and productivity. *American Economic Review, 86*, 263–267.

Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, June–July, pp. 4–16.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay Co., Inc.

Bruce, B. C., & Levin, J. (1997). Educational technology: Media for inquiry, communication, construction, and expression. Retrieved October 8, 1997 from the World Wide Web: http://www.ed.uiuc.edu/facstaff/chip/taxonomy/latest.html

Burge, E. (2000). Using learning technologies: Ideas for keeping one's balance. *Open Praxis*, Vol. 1, pp. 17–20.

Campbell, M. (1999, November 11). Hey, what work? We're cruising the Internet. *The Edmonton Journal*, p. A-3.

Cassidy, J. (2000). The productivity mirage. *The New Yorker*, pp. 106–118.

Chen, Y., & Willits, F. (1998). A path analysis of the concepts in Moore's theory of transactional distance in a videoconferencing environment. *Journal of Distance Education, 13*(2), 51–65.

Chickering, A., & Gamson, Z. (1989). Seven principles for good practice in undergraduate education. *AAHE Bulletin,* March, pp. 3–7.

Clark, R.E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), pp. 445 - 459.

Clark, R.E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-30.

Crooks, B. & Kirkwood, A. (1988). Video-cassettes by design in Open University courses. *Open Learning,* November, pp. 13-17.

Daniel, J. (1996). Implications of the technology adoption life cycle for the use of new media in distance education. In J. Frankl & B. O'Reilly (Eds.), *1996 EDEN conference: Lifelong learning, open learning, distance learning* (pp. 138–141). Poitiers, France: European Distance Education Network.

Dalal, S. (2001, October 26). Futzers draining production budgets. *The Edmonton Journal*, pp. F-1, 8.

Dede, C. (1996). The evolution of distance education: Emerging technologies and distributed learning. *The American Journal of Distance Education, 10*(2), 4–36.

DeSanctis, G., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. *Management Science, 33*(5), 589–609.

Dietrich, J. E., & Johnson, F. C. (1967). A catalytic agent for change in higher education. *Educational Record*, Summer, pp. 206–213.

Evans, J. (1998). Convergances: All together now. *The Computer Paper.* February. Retrieved October 8, 1998 from the World Wide Web: http://www.tcp.ca/1998/9802/9802converge/together/together.html

Fahy, P. J. (1998). Reflections on the *productivity paradox* and distance education technology. *Journal of Distance Education, 13*(2), 66–73.

Fahy, P. J. (2000). Achieving quality with online teaching technologies. Paper presented at the *Quality Learning 2000* Inaugural International Symposium, Calgary, Canada. March. (Available from ERIC documents: ED 439 234.)

Fernandez, B. (1997, October 4). Productivity improvements not computing. *Edmonton Journal*, p. J16.

Fischer, B. (1997). Instructor-led vs. interactive: Not an either/or proposition. *Corporate University Review,* Jan/Feb., pp. 29–30.

Fletcher, J.D. (1992). Individualized systems of instruction. Institute for Defense Analyses.

Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart and Winston.

Gale, S. F. (2001). Use it or lose it. *Online Learning, 5*(7), 34–36.

Garrison, D.R. (1989). *Understanding distance education: A framework for the future.* New York: Routledge.

Girard, K. (2003). Making the world safe for software. *Business 2.0, 4*(5), 64–66.

Gordon, E. E. (1997). Investing in human capital: The case for measuring training ROI. *Corporate University Review, 5*(1), 41–42.

Grabe, C., & Grabe, M. (1996). *Integrating technology for meaningful learning* (pp. 243–247). Toronto: Houghton Mifflin Co. Retrieved February 1999 from the World Wide Web: http://www.quasar.ualberta.ca/edmedia/ETCOMM/readings/Krefgra.html

Greenaway, N. (2002). Internet falling short of hype. *The Edmonton Journal*, June 12, p. A-13.

Grow, G. (1991). Teaching learners to be self-directed. *Adult Education Quarterly, 41*(3), 125–149.

Hartnett, J. (2002). Where have all the Legos gone? *Online Learning, 6*(2), 28–29.

Harvard computing group. (1998). Knowledge management-return on investment. Author. Retrieved March 14, 2000 from the World Wide Web: http://www.harvardcomputing.com

Heinich, R., Molenda, M., Russell, J. D., & Smaldino, S. E. (1996). *Instructional media and technologies for learning* (5th ed.). Englewood Cliffs, NJ: Merrill, an imprint of Prentice Hall.

Helm, P., & McClements, R. (1996). Multimedia business training: The big thing or the next best thing? In J. Frankl, & B. O'Reilly (Eds.). *1996 EDEN conference: Lifelong learning, open learning, distance learning* (pp. 134–137). Poitiers, France: European Distance Education Network.

Howard, B. (2001). 20 years of missed opportunities. *PC Magazine, 20*(15), 75.

Johnson, V. (2000). Using technology to train weather forecasters. *T.H.E. Journal Online*. June. Retrieved March 21, 2002 from the World Wide Web: http://www.thejournal.com/magazine/vault/articleprintversion.cfm? aid=2880

Joyce, B., & Weil, M. (1980). *Models of teaching* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Juler, P. (1990). Promoting interaction; maintaining independence: Swallowing the mixture. *Open Learning,* pp. 24–33.

Kemp, J. E. (1977). *Instructional design* (2nd ed.). Belmont, CA: Fearon-Pitman Publishing.

Khan, S., & Hirata, A. (2001). Lowering the TCO of video communications. Retrieved February 13, 2002 from the World Wide Web: http://www.tmcnet.com/tmcnet/articles/0501en.htm

Koumi, J. (1994). Media comparisons and deployment: a practitioner's view. *British Journal of Educational Technology, 25*(1), pp. 41-57.

Kozma, R. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42*(2), pp. 7 - 19.

Laudon, K., Traver, C., & Laudon, J. (1996). *Information technology and society* (2nd ed.). Toronto: Course Technology Inc.

Levin, R. R., Anglin, G. J., & Carney, R. R. (1987). On empirically validating functions of pictures in prose. In D. M. Willows, & H. A. Houghton (Eds.), *The psychology of illustration: Volume 1, Basic research* (pp. 51–85). New York: Springer-Verlag.

Lohr, S. (1997, October 12). The future came faster in the good old days. *The Edmonton Journal*, p. B-1.

Maier, P., Barnett, L., Warren, A., & Brunner, D. (1996). *Using technology in teaching and learning.* London: Kogan Page.

Marchionini, G. (1988). Hypermedia and learning: Freedom and chaos. *Educational Technology*, pp. 8–12. Retrieved January 1999 from the World Wide Web: www.quasar.ualberta.ca/edmedia/ETCOMM/readings/Krefmar.html

Massy, W. F., & Zemsky, R. (1999). Using information technology to enhance academic productivity. Retrieved October 7, 1999 from the World Wide Web: http://www.educause.ed/nlii/keydocs/massy.html

Matkin, G. (1997). Using financial information in continuing education. Phoenix, AZ: American Council on Education.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

McIsaac, D. (1979). Impact of personal computing on education. *Association for Educational Data Systems Journal, 13*(1), 7–15.

Mehlinger, H. (1996). School reform in the information age. *Phi Delta Kappan*, pp. 400–407.

Mezirow, J., & Irish, G. (1974). Priorities for experimentation and development in adult basic education. *Vol. 1, Planning for innovation in ABE.* New York: Columbia University, Center for Adult Education. (ERIC ED 094 163.)

Miller, M. J. (2002). Broadband optimism. *PC Magazine, 21*(3), 7–8.

Moore, M. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). New York: Routledge.

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education, 3*(2), pp. 1–6. Retrieved November 9, 2001 from the World Wide Web: http://www.ed.psu.edu/acsde/ajde/ed32.asp

Murgatroyd, S. (1992). Business, education, and business education. In M. G. Moore (Ed.), *Distance education for corporate and military training* (pp. 50–63). Readings in distance education, No. 3. University Park, PA: Penn State University, American Center for the Study of Distance Education.

Oberlin, J. L. (1996). The financial mythology of information technology: The new economics. *Cause/Effect*, pp. 21–29.

PC Magazine. (2003c). Broadband: Bringing it home. *PC Magazine*, *22*(5), 25.

Perry, T. (2000). A history of interactive education and training. Retrieved February 4, 2002 from the World Wide Web: http://www.coastal.com/WhatsNew/online_history.html

Picard, J. (1999, June 10). *Creating virtual work teams using IP videoconferencing.* Presentation at the Distance Education Technology '99 Workshop, Edmonton, Alberta.

Quinn, J., & Baily, M. (1994). Information technology: The key to service performance. *Brookings Review, 12*, summer, pp. 36–41.

Rieber, L., & Boyce, M. (1990). The effects of computer animation on adult learning and retrieval tasks. *Journal of Computer-Based Instruction, 17*, pp. 46–52.

Roblyer, M. D., & Schwier, R. A. (2003). *Integrating educational technology into teaching, Canadian edition.* Toronto: Pearson Education Canada Inc.

Roblyer, M. D., Edwards, J., & Havriluk, M. A. (1997). *Integrating technology into teaching* (pp. 27–53). Columbus: Merrill.

Rogers, E. M. (1983). *Communication of innovations* (2nd ed.). New York: The Free Press.

Saettler, P. (1990). *The evolution of American educational technology.* Englewood, CO: Libraries Unlimited, Inc.

Senge, P. (1990). *Fifth discipline*. Toronto: Doubleday.

Stafford, D. (1999, December 15). Surfing from web-linked worksite a common practice, survey shows. *Edmonton Journal*, p. F-7.

Strauss, M. (1997, October 7). Web sites don't boost sales, survey of retailers says. *Globe and Mail*, p. B-8.

Stringer, R. A., & Uchenick, J. (1986). *Strategy traps.* Toronto: Lexington Books.

Szabo, M. (1998). *Survey of educational technology research.* The Educational Technology Professional Development Project (ETPDP) Series. Edmonton, Alberta: Grant MacEwan Community College and Northern Alberta Institute of Technology.

Thaler, J. (1999, May 15). Web in the workplace: Waste or help? *The Edmonton Journal*, p. I-1.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research, 20*(1), 3–43.

Watkins, K., & Callahan, M. (1998). Return on knowledge assets: Rethinking investments in educational technology. *Educational Technology, 38*(4), 33–40.

Welsch, E. (2002). Cautious steps ahead. *Online Learning, 6*(1), 20–24.

Where's the broadband boom? (2002). *PC Magazine, 21*(16), 23.

Wolfe, D. (1990). The management of innovation. In L. Salter, & D. Wolfe (Eds.), *Managing technology* (pp. 63–87). Toronto: Garamond Press.

Workers find online surfing too tempting. (2000, February 22). *The Edmonton Journal*, p. A-3.

Wysocki, B. (1998). Computer backlash hits boardrooms. *The Edmonton Journal*, May 1, p. D-3.

Zimmerman, H. (1972). Task reduction: A basis for curriculum planning and development for adult basic education. In W. M. Brooke (Ed.), *ABE: A resource book of readings* (pp. 334–348). Toronto: New Press.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, pp. 1-24, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.8
E-Learning and Multimedia Databases

Theresa M. Vitolo
Gannon University, USA

Shashidhar Panjala
Gannon University, USA

Jeremy C. Cannell
Gannon University, USA

INTRODUCTION

E-learning covers the variety of teaching and learning approaches, methodologies and technologies supporting synchronous or asynchronous distance education. While distance education is a concept typically used by conventional institutions of education to mean remote access and delivery of instruction, the concept of e-learning broadens the scope to all instances of learning using Web-mediated learning. The scope includes realizing learning organizations (Garvin, 1993), achieving knowledge management (Beccerra-Fernandez; Gonzalez & Sabherwal, 2004; Aussenhofer, 2002) and implementing organizational training.

Individuals continue to learn throughout their lives, particularly as a function of their work and profession. The manner in which they access information and use it often depends upon the available technology, their previously learned response for information acquisition and how their organization facilitates learning and knowledge transfer (Tapscott, 1998; Zemke, Raines & Filipczak, 2000). Hence, e-learning is not simply a consideration for traditional learning institutions, but for any organization.

As such, e-learning not only faces the traditional challenges of teaching to various learning styles while conveying the spectrum of educational objectives, but also faces the extra challenge of using emerging technologies effectively. The three significant emerging technology areas to e-learning are: networking, mobility and multimedia. These technologies can enable a highly interactive delivery of material and communication between instructors and students. Out of the three, however, multimedia technologies relate directly to pedagogical concerns in providing material

tailored to the content domain, to the individual and to the learning objectives (Vitolo, 1993).

Currently, multimedia and e-learning initiatives focus on the presentation of multimedia. The adequate presentation of multimedia is often more an issue of the network being used and its connectivity parameters. Acceptable multimedia presentation depends upon the format of the multimedia and its ability to be quickly transferred (David, 1997). In these circumstances, the availability and appropriateness of the multimedia is assumed to have already been decided as necessary to the instruction.

Not being addressed currently is the storage of multimedia. Multimedia databases should allow for retrieval of components of the integrated and layered elements of the media data stored. In this way, the media would support learning goals. Its retrieval should be conditional upon a context and a content need. Context involves the learning situation – the educational objectives and the learner, combined. Content need includes the particular material to be acquired. Conditional retrieval of multimedia based upon a pedagogical circumstance implies that not all learners or situations need the same media to be delivered, but that a compendium of stored media should be available. In fact, the media alone cannot solely enable learning. Clark (1983) analyzed the effects of learning from different media and observed that significant changes in learning are a function of the media used for the presentation of the material. Significant attention must be given to the content material available for e-learning systems. The material in a certain media format should be included, because it adds or complements the underlying informational intent of the system.

Further, as educational objectives aspire to higher levels of competency such as analysis, synthesis and evaluation, more depth and variety of detail need to be communicated to the student. However, due to the connectivity issues of e-learning, often layers of representation are not available to the learner. For example, during face-to-face communication, student to teacher, the teacher provides the path to the solution and essentially trains the student when teaching analysis skills. However, with e-learning systems, just the end product—the "solution"—of the analysis is provided. When the underlying reasoning layers of the analysis are not available, the overall quality of the instruction suffers (Vitolo, 2003).

Multimedia databases added to an e-learning initiative would provide conditional retrieval and comprehensive storage of multimedia. However, no database management system (DBMS) exists solely for multimedia storage and access (Elmasri & Navathe, 2000). Several current DBMS do provide a data type appropriate for multimedia objects. However, the range of capabilities available for manipulating the stored object is severely limited. A pure multimedia database management system (MDDBMS) is not commercially available, now.

BACKGROUND

Learning, education and teaching are inextricably intertwined, highly complex processes. Each process has been researched as a social phenomenon, cognitive transformation, generational bias and personality expression. While the work on these topics is vast, several aspects are generally accepted as foundation concepts:

- People interact with environments on an individualized basis. Learners have learning styles; teachers have teaching styles; individuals have personality styles.
- Educational efforts seek to find a correspondence between these various styles so that learning can progress effectively.
- Educational efforts can be described via taxonomies—progressions of objectives. The realization of these objectives does not necessarily require any specific learning or

teaching modality. The communication of the content of the objective may be better suited to one modality (visual, auditory or tactile) than another.
- Learning can continue throughout an individual's life.
- Technology can facilitate educational efforts by providing various formatted and comprehensive content for interactive and self-regulated learning. Multimedia technology provides an excellent opportunity for packaging content into a variety of modalities.

With respect to styles, Coates (2002) provides a condensation of the various style-based perspectives of learning. While much of these style-based analyses of behavior stem from the initial work of Carl Jung (1923), the facets of the styles are continually being researched. Learning is mediated by a variety of factors—some (such as modality of instruction) that can be manipulated successfully within an educational effort, some (such as generational cohort biases) that are out of the control of instructional design.

With respect to educational structures, educational researchers have developed taxonomies to explain educational objectives. (See Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths & Wittrock, (2001) and Bloom (1984, 1956) for classic coverage of these taxonomies.) Essentially, educational efforts advance instruction in levels of difficulty and performance so that the breadth and depth of the knowledge of a field can be communicated.

As a foundation concept to using multimedia for e-learning, the media requires appropriate processing for adequate capture, production and distribution. For example, video may be shot using either an analog or digital camera. Before the source video can be edited using computer software, it must be instantly accessible from a hard disk and not the original videotape. The source video is imported into the computer by a process called video capture. Captured video is huge; 10 seconds of raw, uncompressed NTSC video (the standard for television video) use as much as 300 megabytes (MB) of storage space.

For video to be played in a Web browser or distributed on CD-ROM, the file size must be reduced significantly. This file size reduction, or compression, is achieved using codecs – compression/ decompression approaches. Source video captured from a digital camcorder will already have been digitized and saved in a digital file format inside the camera. Digitizing a video sequence results in extremely high data rates. For example, an image with a resolution of 720x576 pixels and a color depth of 16 bits produces a data stream of 1.35 MB per individual frame. At the rate of 25 frames per second required to render smooth video scenes, a gigantic data volume of 3,375 MB/second results. This volume is far too great for the average hard disk to handle; a CD-ROM would only have enough space for about 16 seconds (Adobe Press, 2003; Bolante, 2004).

Next, the capture process involves transferring the digitized video file to a computer hard disk. Once captured, the multimedia requires further considerations for production and dissemination considerations. The analog or digital source video is captured using video editing software and saved into an appropriate video format. These video formatted files are also large; 60 minutes of video can consume 12 GB of disk space. The media file is manipulated within software via timing option, making it ready for rendering and production. After rendering, the video file is processed further depending upon its desired distribution modality:

- Exported back to video tape (analog or digital)
- Compressed further for distribution on CD-ROM or DVD
- Compressed further for distribution across the Internet

The final presentation also has options. Progressive encoding refers to where the entire video must be downloaded before any viewing occurs, regardless of its format. This case occurs with any of the formats considered so far. Alternatively, Internet streaming enables the viewer to watch sections of video without downloading the entire file. Here, the video starts after just a few seconds. The quality of streaming formats is significantly lower than progressive formats due to the compression being used (Menin, 2002).

Finally, appropriate display of the material for effective consumption is improved with interactive multimedia. However, interaction with a media file—the goal of interactive multimedia—is restricted; navigation is possible using pause, forward and reverse controls provided by the player installed on the client computer. To create interactive media for the Web, CDs, kiosks, presentations and corporate intranets, a multimedia authoring program is used. These programs enable the combination of text, graphics, sound, video or vector graphics in any sequence. To add more interactive features, powerful scripting languages are also provided (Gross, 2003).

Hence, the situation for e-learning is bound in several ways by the available multimedia technology. First, the production and distribution of multimedia is not a trivial undertaking, requiring specialized skills and technologies. Second, the viewing of the multimedia requires the client machine and user to have appropriate technology. Third, the goals of the e-learning effort must be in balance with the available and expected technology. Fourth, the multimedia technology itself is providing limited options for interactive manipulations. After these steps, the media as a data-rich structure can be stored in multimedia databases.

LIMITED STATE OF E-LEARNING AND MULTIMEDIA DATABASE SYSTEMS

The requirements for the next era of e-learning applications using multimedia databases are:

1. Repository systems offering storage and access capabilities of media
2. Indexable storage structure for media files as contiguous structures composed of identifiable and searchable elements
3. Tier-architecture deployment providing multiple application access

With respect to point 1, DBMS implementations are commercially available that can reference media files in a variety of formats. The media file is handled as a complete unit through a large object (LOB) data type. For instance, Oracle introduced a set of LOB data types with Oracle 8 to facilitate the storage of large-scale digitized structures and references to them. The media itself is stored in one of two ways: as a LOB (usually BLOB) type within the database, or in an external file and pointed to by a BFILE type within the database.

Oracle has continued to advance the integration of these LOB data types through its database versions and the various tools it offers. Oracle's *inter*Media management system and Oracle 10g database support various media specific object types, recognize and record facets of the media's attributes in metadata structures, and provide support to multimedia needs for various applications and enterprise-wide delivery. The *inter*Media objects of ORDAudio, ORDImage, ORDVideo and ORDDoc provide attributes and access capabilities recognized by the end-user application (Oracle, 2003, 1999).

With respect to point 2, however, media data are not handled as an indexed structure. Thus, access to incremental slices or partitions of the media is not a current capability. Access to the

attributes of the entire media as a unit has been improved, but more is desired for e-learning. A beginning point would be a query standard for multimedia and its internal elements. Oracle 10g does support a portion of the first edition of the ISO/IEC 13249-5:2001SQL/MM Part 5: Still Image Standard. The standards community and commercial vendors continue to address the query needs of multimedia and applications.

With respect to point 3, multimedia applications may not have any database component. The media is handled as a data item manipulated by the script of the application (David, 1997). This situation mimics the file-processing era; namely, that the data manipulated by one file for its application's needs may not lend itself to manipulation for another, separate application's needs.

A more desirable architecture for multimedia delivery is a three-tier one. The advantages of a multi-tier architecture are:

- Separation of the user interface logic and business logic
- Low bandwidth network requirements
- Business logic resides on a small number (possibly only one) of the middle machines

Together, these aspects promote greater accessibility across various applications and ease of maintenance through hardware and software upgrades.

E-learning has benefited from the enhanced broadband and accessibility of the public infrastructure. As such, the popularity and market presence of the initiatives have grown. The initiatives offer delivery convenience, communication channels and content volume. These three factors make e-learning a highly attractive possibility for a variety of learning circumstances.

In many respects, however, current e-learning initiatives are similar to page-turning, computer-assisted instruction packages of earlier computer-based learning efforts. That is, the initiatives lack pedagogical development and refinements, tailoring the instruction with respect to a student model and to the complexities of higher-level educational objectives.

The higher levels of educational objectives need to communicate greater complexity in the detail, explanation and incremental refinements of the content. If only the final result of the analysis, synthesis or evaluation is the goal of the instruction, then current multimedia efforts would be fine. However, higher-level educational objectives require more layers and connections to be communicated—more complex structures need to be communicated in order to teach more complex concepts. Further, this type of refinement of complex structures through incremental layers of development, feedback, and progress is desirable for effective education and for effective handling of learning styles given a learning situation. To achieve mediated learning episodes with this finesse, the next generation of multimedia applications and deployment utilities are necessary.

FUTURE TRENDS & CHALLENGES

Multimedia capabilities will continue to improve, becoming more economical and more usable. In time, the authoring, production and distribution of multimedia will become as easy as word processing. As with many information systems efforts, the challenge resides with understanding the infrastructure commitment to deploy such efforts in terms of hardware, skills and procedures. Successful efforts require high-capacity, secure servers and connections. Individuals need to understand the nature of multimedia to manipulate it successfully within the software. Finally, well-defined procedures for the distribution and maintenance of the multimedia over a desired architecture must accompany the effort and must be handled by systems staff cognizant of the desired performance levels.

The future of multimedia databases shares this same positive outlook. The capabilities sought in various data-typing of media, querying of media segments and indexed aspects need continued addressing.

For e-learning, the challenges parallel those of multimedia. E-learning efforts need to understand how the infrastructure can limit the instructional goals. The skill level for development efforts requires technical competence and instructional design principles. When future e-learning efforts include multimedia databases, then the required technical skills will be further specialized. E-learning efforts will require teams of highly specialized individuals, bridging the different technical needs for pedagogy, multimedia and multimedia databases.

The final future challenge to be addressed is one shared with many Web-based developments—intellectual property rights. Intellectual property is a sufficiently difficult concept currently when multimedia is part of a single application. Once the multimedia is part of applications connected through a database, then the intellectual property rights of the database and its development must be considered also.

CONCLUSION

E-learning continues the efforts of computer-mediated instruction. The depth and interactivity potential of multimedia components is a highly attractive factor to add to instruction. Multimedia offers the capability to construct interactions tailored to the learning needs of a specific student, within a specific learning context, being taught a specific content domain.

Multimedia technology has matured significantly as its complementary technologies of network capacities and deployment hardware have advanced. However, for the next generation of multimedia and e-learning to progress, multimedia databases should be used. The database configuration would increase the potential use of the multimedia across multiple application instances, the multimedia could be queried for access and, ultimately, the elements composing the multimedia could be accessed as opposed to accessing the entire multimedia file—the current option for multimedia access.

Multimedia databases not only would enhance the technical delivery of e-learning efforts, but also would enhance the pedagogical aims of e-learning efforts. Instruction of the higher-order educational objectives requires layers of a representation to be presented. Multimedia databases could store media in its elemental segments so that selective delivery of pieces of the media could be offered for instruction—not the media file in its entirety, leaving the parsing of the relevancy of the media to the discretion of the student.

While multimedia databases would increase the flexibility, access and reuse of the media, other challenges arise. Multimedia databases are complex technologies requiring more specialized skills beyond simply building and deploying multimedia. Adequately supporting multimedia databases requires continued, expensive investments in infrastructure to support the deployment of the databases and e-learning efforts. Further, not all of the required features of true multimedia databases have been developed to date, but are part of the current efforts of database developers and of standards communities. Finally, intellectual property issues are a challenge of applications using multimedia databases. As in most development aspects, the intellectual property issues will be as difficult to resolve as the technology was to develop.

REFERENCES

Adobe Press. (2003). *Adobe Premiere 6.5: Classroom in a book*. San Jose: Adobe Systems International.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing.* New York: Longman Publishers.

Ausserhofer, A. (2002). E-learning & knowledge management towards life-long education. Graz, Austria: Competence Center for Knowledge-based Applications and Systems. Retrieved December 10, 2003, from *www.know-center.tugraz.at/de/divisions/publications/pdf/aausser2002-01.pdf*

Becerra-Fernandez, I., Gonzalez, A., & Sabherwal, R. (2004). *Knowledge management: Challenges, solutions, and technologies.* Upper Saddle River, NJ: Pearson Education.

Bloom, B.S. (Ed.) (1984, 1956). *Taxonomy of educational objectives. Handbook 1: Cognitive domain.* New York: Longman.

Bolante, A. (2004). *Premiere Pro for Windows.* San Jose: Peachpit Press.

Clark, R.E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Coates, J. (2002). *Generational learning style*. Retrieved May 23, 2002, from *www.lern.org/gls_booklet.pdf*

David, M.M. (1997). Multimedia databases through the looking glass. *Intelligent Enterprise's Database Programming & Design: On-Line*. Retrieved December 10, 2003, from *www.dbpd.com/ vault/9705davd.htm*

Garvin, D.A. (1993). Building a learning organization. *Harvard Business Review, 71*(4), 78-92.

Gross, P. (2003). *Macromedia Director MX and lingo: Training from the source*. Berkeley, CA: Macromedia Press.

Jung, C. (1923). *Psychological types.* New York: Harcourt & Brace.

Menin, E. (2002). *The streaming media handbook*. Upper Saddle River, NJ: Prentice Hall.

Navathe, S.B., & Elmasri, R. (2000). *Fundamentals of database systems* (3rd ed.). Reading, MA: Addison-Wesley.

Oracle. (1999). *Using Oracle 8i* inter*Media with the Web, Release 8.1.5.2* (Part No. A77033-01). Retrieved May 20, 2004.

Oracle. (2003). *Oracle* inter*Media: Managing Multimedia Content* (Oracle white paper). Retrieved May 20, 2004.

Tapscott, D. (1998). *Growing up digital: The rise of the Net generation*. New York: McGraw-Hill.

Vitolo, T.M. (1993). The case for self-revealing multimedia systems. *Proceedings of the 11th Annual Conference of the Association of Management,* Atlanta, Georgia, August 5-9.

Vitolo, T.M. (2003). *The importance of the path not taken: The value of sharing process as well as product. Final report.* Vancouver: SMARTer Kids Foundation.

Zemke, R., Raines, C., & Filipczak, B. (2000). *Generations at work: Managing the clash of veterans, boomers, Xers, and nexters in your workplace*. New York: AMACOM.

KEY TERMS

Bit Depth: The number of bits used for color resolution when viewing a movie.

Codec: Compression and decompression algorithms provided by either a software application or a hardware device.

Database Management System (DBMS): Collection of software components to store data, access the data, define data elements, store data element definitions, build data storage structures, query the data, backup and secure the data, and provide reports of the data.

E-Learning: All teaching and learning processes and functions from course authoring, course management, examinations, content delivery, feedback and course administration developed, delivered and monitored through synchronous or asynchronous communication.

Encoding: The process of using codecs to convert video files to different distribution file formats. The codecs used for encoding files for CD-ROM and DVD are MPEG-1 and MPEG-2, respectively.

Frame Rate: The number of frames projected per second.

Frame size: The height and width of the video window according to the number of pixels.

Internet Streaming: Video format that intermittently downloads sections of a media file to a client.

Knowledge Management: The set of initiatives to identifying, retrieving, organizing, disseminating and leveraging intellectual capital—usually as an enterprise-wide effort.

Learning Styles: A cognitive perspective of individualized preferences for modalities when learning; includes a learner's manner of responding to and using stimuli while learning.

Multimedia Database: Database storage and retrieval capabilities developed with respect to multimedia requirements for high-quality, rapid, queried usage by applications.

Progressive Encoding: Video format that downloads the entire media file to the client before any displaying occurs.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 271-277, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.9
Integrating Multimedia Cues in E-Learning Documents for Enhanced Learning

Ankush Mittal
Indian Institute of Technology, India

Krishnan V. Pagalthivarthi
Indian Institute of Technology, India

Edward Altman
Institute for Infocomm Research, Singapore

INTRODUCTION

The digitization of educational content is radically transforming the learning environment of the student. A single lecture, as well as supporting reference material, textbook chapters, simulations, and threaded chat system archives, may be captured on one hour of video, a set of 20 or more slides, and ancillary text resources. A single course may contain 25 such lectures and a single department may have 30 distinct courses that have been digitized in a single year. If, while taking one course, a student wants to find a relevant definition, example, discussion, or illustration of a concept, then the student would potentially need to search as much as 750 hours of video, 15,000 slides, and a huge volume of text in order to find the desired information. Thus the online student is overwhelmed by a flood of multimedia data which inhibits the development of insight.

Insight is a key ingredient of education and is most often achieved by the manipulation of information through the discovery of new relationships, identification of hidden structures, or the construction of domain models. The methodology of instructional design may be used to anticipate the needs of the student in the controlled environment of a classroom with novice learners, but it is inadequate for the needs of a heterogeneous population of online learners. Intelligent tutoring systems provide an additional degree of flexibility for the independent learner, but are

difficult to produce and maintain. The advanced learner needs powerful search and organizational tools to support self-guided learning. In all three cases there is a need for content-based retrieval of multimedia resources ranging from simple indexing and navigation of lectures to ontology-based mining of information nuggets from large repositories of heterogeneous content.

The needs of the independent learner are particularly demanding due to the requirements for real-time, context-dependant, and precise retrieval of unstructured and incomplete information distributed across multiple media sources. Due to the large size of the corpus and the highly focused nature of the target information, the systematic labeling (either manually or automatically) of the media is not feasible. Instead, we propose a system for *media hot spotting.*

Media hot spotting is the process of finding *hot spots* within the text, audio, video, and other media content. A hot spot may be a distribution of key terms in a text document; a matching distribution of terms in the speech track of a lecture video; or a sequence of writing, emphasis, and gesture events in the video stream of a lecture, where an event is a spatial or temporal set of interrelated features. Individually, each piece of information does not convey sufficient semantic information to identify the informational content of the media. In combination, they provide significant evidence, for instance, that a Definition event has occurred in the media and that a particular term has most likely been defined within the context of this event. Thus, given a model for how information co-occurs across different media objects, hot spotting enables the rapid retrieval of candidate media content for further analysis and reference.

Content-based solutions are available for domains like sports and news, but have not yet been systematically explored for educational videos (Idris & Panchanathan, 1997; Mittal & Cheong, 2003; Woudstra et al., 1998). This chapter describes a new method of content-based retrieval for e-learning videos using camera motion cues, audio features, slide layout, and associated decision rules. Using this technique, we are able to separate the lecture videos into several component states and personalize the video from these states. For our experiments, we used 26 lecture videos from the Singapore-MIT Alliance, along with the associated PowerPoint slides.

This chapter is organized as follows. The next section presents a discussion of an existing distance learning program (SMA) and other related work. Then an overview of our approach is presented, followed by the section titled "Multimedia Indexing Features," where we show the framework for modeling multimedia information and present a list of the most useful features used in the video segmentation task. The section titled "Indexing of Lectures" elaborates upon the mapping of low-level features to lecture semantics. Finally, we discuss the experimental results and significance of this approach.

DISTANCE LEARNING PARADIGM

Singapore: MIT Alliance Educational Setup

The work presented here relates to the materials used in the Singapore-MIT Alliance (SMA)[1] development program. SMA is an innovative engineering education and research collaboration among the National University of Singapore (NUS), Nanyang Technological University (NTU), and the Massachusetts Institute of Technology (MIT). SMA classes are held in specially equipped classrooms at the Singaporean institutes and at MIT using live video transmission over the Internet. The synchronous transmission allows participants at both locations to see each other and speak normally. However, because of the 12-hour time zone difference, SMA has made a great effort to find and develop tools to enhance asynchronous learning.

SMA lectures are given daily, and it is expensive to process, index, and label them through manual methods. Immediate availability of the lectures is also important because the courses are fast paced and build sequentially upon earlier lectures. Techniques for the efficient indexing and retrieval of lecture videos are required in order to cope with the volume of lecture video data produced by the SMA program and other media-intensive programs.

SMA is just one example of how the application of information technology to digital media is rapidly enhancing the field of distance education. The effective use of information technology enables institutions to go beyond the classroom to create personalized, lifelong learning for the student. This results in a repository of potential learning experiences that is available not only for the student, but also available for incremental refinement and elaboration by the lecturer.

Related Work

Using video for educational purposes is a topic that has been addressed at least since the 1970s (Chambers & Specher, 1980; Michalopoulos, 1976). Recently the focus of the research has been on the maximum utilization of educational video data which has accumulated over a period of time. Ip and Chan (1998) use the lecture notes along with Optical Character Recognition (OCR) techniques to synchronize the video with the text. A hierarchical index is formed by analyzing the original lecture text to extract different levels of headings. An underlying assumption is made that the slides are organized as a hierarchy of topics, which is not always the case. Many slides may have titles which are in no way related to the previous slide.

Bibiloni and Galli (1996) proposed a system using a human intermediary (the teacher) as an interpreter to manually index the video. Although this system indexes the video, it is highly dependent upon the vocabulary used by the teacher, which may differ from person to person. Moreover, even the same person may have a different interpretation of the same image or video at different times, as pointed out by Ip and Chan (1998).

Hwang, Youn, Deshpande, and Sun (1997) propose a hypervideo editor tool to allow the instructor to mark various portions of the class video and create the corresponding hyperlinks and multimedia features to facilitate the students' access to these prerecorded sequences through a Web browser. This scheme also requires a human intermediary and thus is not generalized.

The recently developed COVA system (Cha & Chung, 2001) offers browsing and querying in a lecture database; however, it constructs the lecture index using a digital textbook and neglects other sources of information such as audio or PowerPoint slides. A similar work (Mittal, Dixit, Maheshwari, & Sung, 2003) concentrated on deriving semantic relationships between concepts and answering queries solely based on PowerPoint slides. Thus, it can be concluded that the integration of the information present in various e-learning materials has not been systematically explored.

STATE MODEL FOR LECTURES AND OVERVIEW OF OUR APPROACH

The integration of information contained in e-learning materials depends upon the creation of a unifying index that can be applied across information sources. In content-based retrieval systems, it is often convenient to create a *state model* in which nodes represent semantically meaningful states and the links between nodes represent the transition probabilities between states. Thus, the methodology for constructing an educational video information system begins with the creation of a state model for the lecture, where the states are based on the pedagogical style of teaching. For the purpose of illustrating the concept, let us consider computer science courses, especially

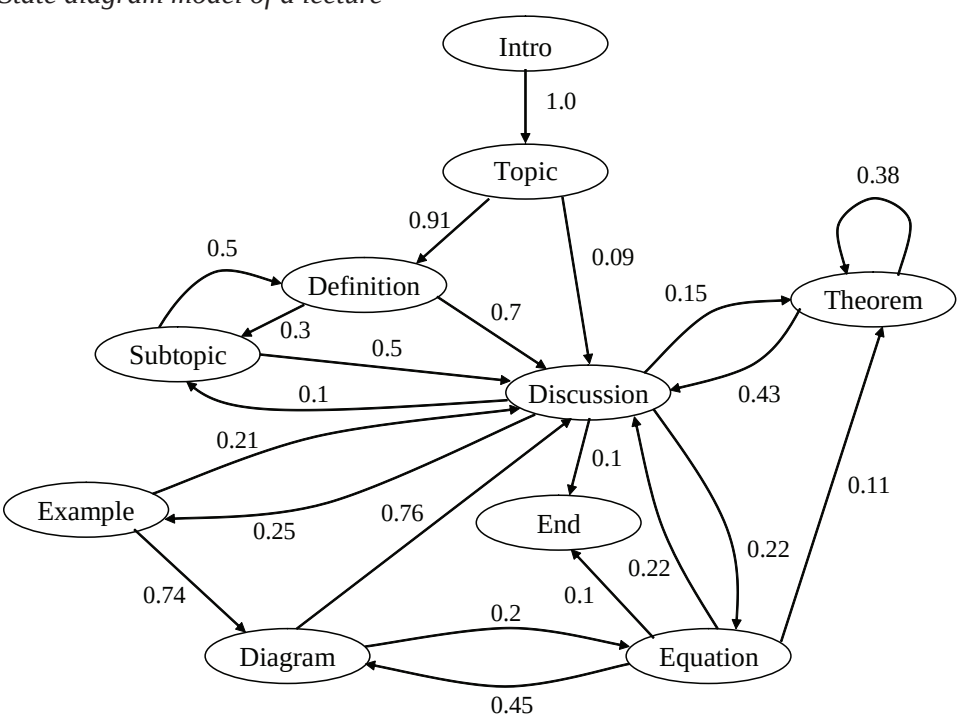
theoretical ones like Introduction to Algorithms. In this case, each lecture can be said to contain one or more topics. Each topic contains zero or more of the following:

- **Introduction:** general overview of the topic
- **Definitions and Theorems:** formal statement of core elements of the topic
- **Theory:** derivations with equations and diagrams
- **Discussions:** examples with equations and diagrams
- **Review:** repetition of key ideas
- **Question and Answer:** dialogue session with the students
- **Sub-Topic:** branch to a related topic

A simple state model for a video-based lecture can be represented as shown in Figure 1. This model for indexing e-learning videos is a state model consisting of 10 different states linked by maximal probability edges. Each state follows the probabilistic edges to go to another state. For example, from state Topic, one potential next state is Definition with a probability of 0.91 and another potential next state is Discussion with a probability of 0.09. Figure 1 shows the corresponding probabilities of transitions from one state to another based on our analysis of the SMA lecture corpus.

The state model implicitly encodes information about the temporal relationships between events. For instance, it is clear from Figure 1 that the introduction of a topic is never followed by a theorem without giving a definition. The state model, supplemented with our indexing techniques (discussed in later sections), provides useful information regarding the possible progression of topics in the lecture. For example, if a lecture is in the Discussion state and has some special attribute such as a camera zoom on the blackboard,

Figure 1. State diagram model of a lecture



we can say with a high level of confidence that it will go next to the Example state. Thus, the state diagram provides a framework for making inferences about lectures which can be used to determine the probable flow of pedagogical transitions that occur in the lecture.

Overview of Our Approach

Various camera motion techniques such as zooming in, zooming out, and panning can be extracted from educational videos. We match sequences of such camera motions to various states in the lectures, such as Theorem, Equation, and so forth (see Figure 1). In addition, we use the information available through the nonvisual media, including text from PowerPoint slides and audio from the lecture, to correctly correlate the heterogeneous media sources. Lastly, we are able to perform contextual searches for media content with minimal use of manual annotations.

The semantic analysis of raw video consists of four steps:

1. **Extract low- and mid-level features:** Examples of low-level features are color, motion, and italicized text. Some mid-level features are zoom-in and increased hand movement of the lecturer. Numerical representations of these features for a particular video segment are then assembled into a vector for subsequent processing.
2. **Classify the feature vectors into a finite set of states within the lecture:** States may correspond to Definitions, Emphasis, Topic Change, Q&A, Review, and so on. These are the objects and events identified in the lecture which are likely to be associated with semantically meaningful events.
3. **Apply contextual information to the sequence of states to determine higher-level semantic events:** such as defining a new term, reviewing a topic, or engaging in off-topic discussion.
4. **Apply a set of high-level constraints to the sequences of semantic events:** to improve the consistency of the final labeling.

Multimedia Feature Extraction and Analysis

Feature extraction is a widely accepted initial step towards video indexing. Features are interpretation-independent characteristics that are computationally derived from the media. Examples are pitch and noise for audio, and color histogram and shape for images. Experience across a wide range of multimedia applications has resulted in the identification of a large number of features that are useful for indexing (Gonzalez & Woods, 1992). The next step is to add semantics to the collection of features (e.g., high pitch in audio, pointing gestures, hand velocity, etc.) so one can use relationships among the features to infer the occurrence of higher level events (Gudivada & Raghavan, 1995).

MULTIMEDIA INDEXING FEATURES

The key influencing factor for the success of any video indexing algorithm is the type of features employed for the analysis. Many features have been proposed for this purpose (Idris & Panchanathan, 1997). Some are task specific, while others are more general and can be useful for a variety of applications.

Audio Features

There are many features that can be used to characterize audio signals. The three features of volume, spoken word rate, and spectral components have proven to be useful for lecture analysis. Volume is a reliable indicator for detecting silence, which may help to segment an audio sequence and determine event boundaries. The temporal variation in volume can reflect the scene content. For

example, a sudden increase in volume may indicate a transition to a new topic. Spoken word rate and recurrence of a particular word are indicators for the scope of a topic or cluster of words within a discussion (Witbrock & Hauptmann, 1997). Finally, spectral component features refer to the Fourier Transform of the samples of the audio signal. Analysis of the frequency components of audio signals using signal processing methods provides support for speaker change detection and other advanced tasks.

Video Features

A great amount of research has gone into summarizing and reviewing various features useful for video segmentation (Wang, Liu, & Huang, 2000). The color histogram, which represents the color distribution in an image, is one of the most widely used color features. The simplest histogram method computes the gray level or color histogram of two frames. If the difference between the two histograms is above the threshold, a boundary shot is assumed.

Motion is also an important attribute of video. Motion information can be generated by block matching or optical flow techniques (Akutsu, Tonomura, Ohba, & Hashimoto, 1992). Motion features such as motion field, motion histogram, or global motion parameters can be extracted from motion vectors. Other video features include texture, compression, and shape, which have been addressed in many papers. The features presented above have been used in different image retrieval systems to effectively perform a video segmentation.

Text Features

In the distance learning paradigm, text is one of the most important features that still has not been researched and utilized extensively. Ip and Chan (1998) propose text-assisted video content extraction, but only to synchronize the video with the text. Text extracted from PowerPoint slides, which are generally provided with educational videos, inherently stores a great deal of information, as we shall see with the SMA lectures.

INDEXING OF LECTURES

The most important and basic steps in a video indexing engine are to extract salient features and then combine these to get the most efficient indexes. A potentially rich source of pedagogical information is the blackboard activity during the presentation of the lecture, which in turn is highly correlated with the content of the lecture notes. Thus, a proper analysis of the lecture notes, which occur as PowerPoint slides, along with the properties discussed below provide an effective means for identifying pedagogical structures in the lecture video.

The PowerPoint slides that are distributed as lecture notes inherently store important information regarding the lecture, which is still largely untapped. Information in the form of the size, shape, color, and boldness of fonts reveals important aspects of the lecture. The state model for the educational videos discussed before enables us to divide the full lecture into four basic categories, namely: Definitions & Theorems, Examples, Proofs, and Formulae. In the analysis of the SMA lectures, we found that there exist stylistic conventions in the PowerPoint slides such as: all the important words (keywords with definitions) are in red and italicized; the special names are always in quotes; the slides having an example always have the word “example” in it; the questions and FAQs have a question mark in the particular slide; and the common names are always in square brackets. Some video features such as camera *zoom in* or *zoom out* to the blackboard or to the audience also specify a transition in the lecture from one state to another state, say

from Diagram to Discussion state. The rules for indexing the slides in the above-mentioned four categories can be summarized as follows.

Category 1: Definitions and Theorems

The keywords or defined terms are always red and in italics, so if there is a definition in the slide, it should have a red italicized word. The word *definition* or *theorem* may be present, but the string queried (the string is the set of words that one is searching for) must be found in the slide.

Category 2: Examples

The course material under consideration for Introduction to Algorithms has an associated image file for all the examples to represent special graphics or equations. The presence of the text pattern '*examples*' or '*examples:*' along with the string queried, is mandatory for a slide to qualify as one containing examples.

When analyzing the text, the context of the current slide is related to the previous slides. Figure 2 illustrates a case where a single example is presented in a sequence of slides. The sequence of processing steps for the well-known sorting algorithm called merge sort is illustrated in these slides. Progressive changes between consecutive slides provide evidence for labeling the associated video as the Example state. In the case of Figure 3, there is an example embedded in a definition as indicated by the word *example* found towards the bottom of the slide. Thus, the context of the particular example is linked to the contents above it and the topic currently being discussed.

Category 3: Proof

The word *proof* along with the string queried is assumed to be present in the slides having relevant information associated with the query. This assumption is a generalized one and can be used for all distance courseware.

Category 4: Formulae

Slides containing embedded formulae can be easily identified through the identification of special symbols used to represent the mathematical expressions. Queries for mathematical expressions can be resolved by converting the query expression into a string of characters and then performing

Figure 2. Slides illustrating the determination of context on the basis of text

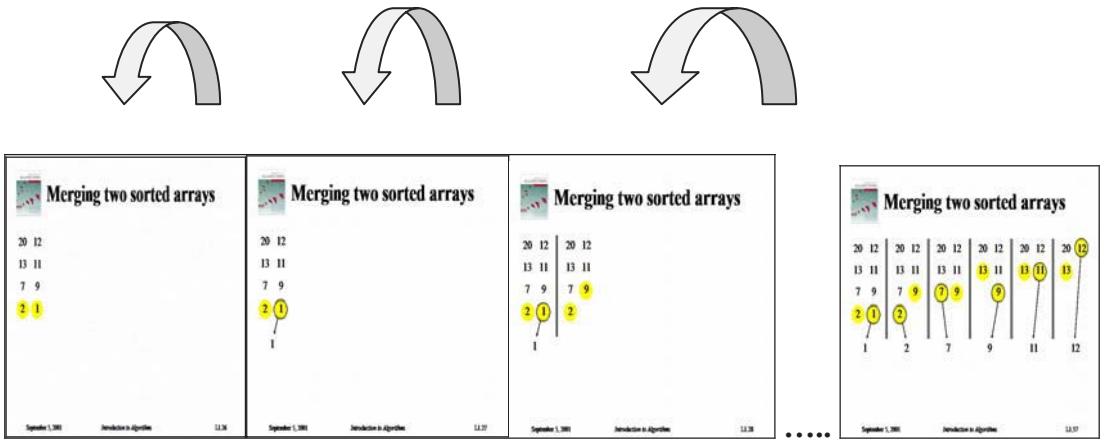
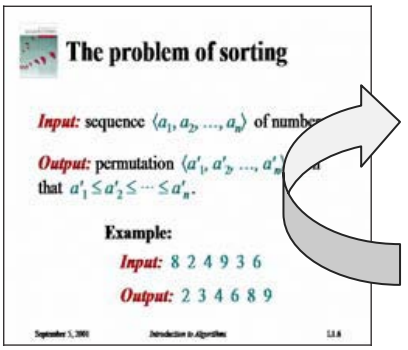


Figure 3. This slide shows the definition of a concept with an embedded example



pattern matching. Special attention must be given to matching expressions that may assume alternate forms, such as fractions and summations.

This set of rules for indexing the slides is supplemented by video features to improve the overall quality of the video indexing. It is well known that video features taken in isolation are highly ambiguous. The video feature for *zoom out* may indicate either discussion in the class, the presentation of a definition, or simply a technique to reduce the tedium of the video (see Figure 4). Similarly, we find that the video feature *zoom in* may indicate the occurrence of the Topic, Example, Diagram, Theorem, or Equation states. Although there is a large overlap between pairs of features, the combination of multiple features given in Table 1 dramatically improves the accuracy of the state classification.

After the entire lecture has been classified, the labeled metadata can be used to perform multiple functions. The first one is searching in context. Several automatic frameworks exist for searching in context (e.g., Mittal and Altman, 2003). Here we employ a simple contextual searching algorithm. To enable searching in context, we need to manually enter the topic names for each video clip associated with a significant pedagogical event identified by the application of the classification rule set. Once the topic names have been keyed into the topic lists, we can then perform a contextual search just by searching for all occurrences of the queried subject and returning the results. This method is accurate because under our definition of the Topic state, all subject matter that is important enough to be explained separately is classified as a Topic or a Subtopic. For example, when *quicksort* is mentioned under the divide-and-conquer method (divide-and-conquer is a generic technique and quicksort is its specific application), our system classifies *quicksort* as a subtopic. Again, when *insertion sort* is compared with *quicksort*, it classifies *insertion sort* as another subtopic. As a result, the topic list is comprehensive in covering all material that is of importance. Hence, we are able to retrieve all instances of a particular query by searching through the topic list.

Figure 4. (a) Zoom in to the lecturer; (b) Zoom out to the room

(a)

(b)

Table 1. Summary of the video and text features

State	Video Features	Text Features
Topic	1) Zoom In 2) Stay Zoomed In 3) Underline on board	1) Slide Title
Definition	1) Zoom Out 2) Input and Output on board	1) Defined Word is Red Italicized 2) Presence of word Definition or Theorem along with defined word
Example	1) Zoom In 2) Ex: on blackboard	1.) Presence of word Example: or Examples along with topic 2) An associated *.gif file
Discussion	1) Zoom out for entire class 2) Change in voice OR 1) Zoom In on lecturer 2) Increased hand movement of lecturer	1) No or much less blackboard activity
Theorem	1) Zoom In 2) Theorem or Proof or Corollary in PPT 3) Same as (2) but on blackboard	1) Defined Word is Red Italicized 2) Presence of word Definition or Theorem
Formulae & Equation	1) Zoom In on blackboard	1) Associated *.gif file 2) Associated *.wmf file

EXPERIMENTAL RESULTS AND APPLICATIONS

The key idea is to create a system that automatically segments the educational videos so students can then use it to view the desired sections of the lectures without going through a linear search, thereby saving them time and effort. We tested our method on 26 lecture videos from the Singapore-MIT Alliance course SMA5503. The semiautomatic classification results are tabulated in Table 2.

Overall, our method has an accuracy of 85.1% in detecting the correct state. The personalization rules dependent on the first algorithm also have an accuracy of 85.1%. The contextual search algorithm is solely dependent on the correct clas-

Table 2. Experimental results for the detection of the lecture states

TOPIC	**DETECTION ACCURACY (%)**
Introduction	100
Topic	90
Definition	80
Discussion	86
Theorem	87.5
Example	83
Equation	62
Diagram	92.3

sification of the topic state and has an accuracy of 90%.

Further, we are able to efficiently and accurately search in context throughout the video database. For example, by searching for the string *merge sort*, we return not only the video clip that teaches *merge sort*, but also other clips from other lectures where some aspect of *merge sort* is further explained (see Figure 5). In this particular case, *merge sort* is mentioned in video lecture 1 under the topic Sorting. It is also mentioned again in lecture 3 under Time Analysis, and in lecture 7 where it is compared to *quicksort*. Hence, when a student uses this system to search for *merge sort*, he has immediate access to all three related video clips even though they are taught in completely different lectures and different parts of the course. As a result, a student searching for *merge sort* will get a much clearer idea of how it actually works and all its different aspects. The combination of semiautomatic analysis of lecture videos to identify pedagogical events and key topics along with techniques for discovering emergent semantics from the media metadata provides the student valuable insights into the finer concepts of the queried subject. By using a simple user interface, users can enter keywords, view topic associations, and search for related materials as shown in Figure 6 (Altman & Wyse, 2005).

CONCLUSION

This chapter presents a system for indexing videos using audio, video, and PowerPoint slides and then segmenting the video content into various

Figure 5. Search for Merge Sort

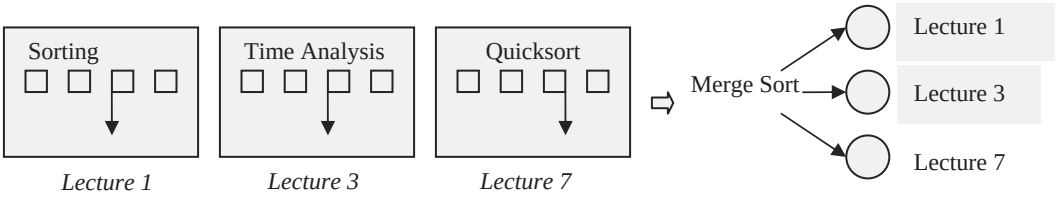
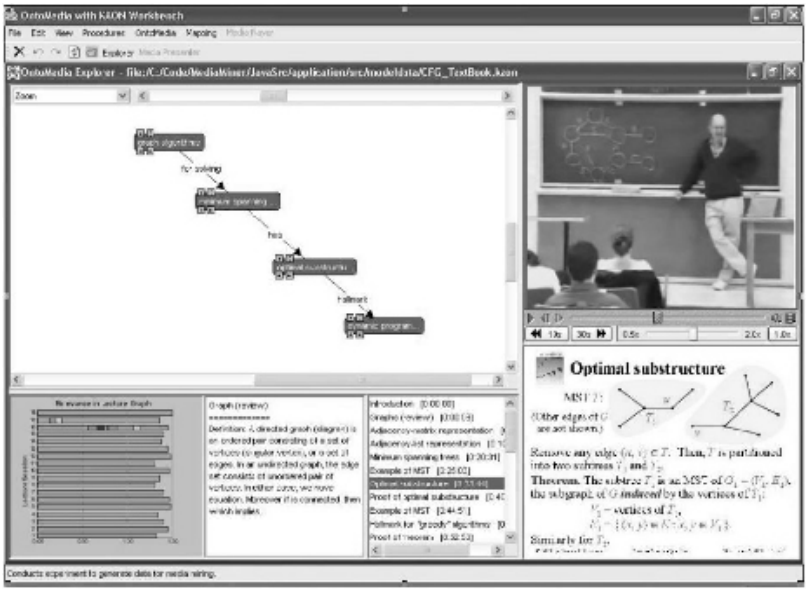


Figure 6. Graphical user interface (GUI) for the presentation of lecture content in response to a query



lecture components. The use of multiple media sources ensures accuracy in the classification of lecture content. This is of critical importance for the effective support of e-learning tasks and maintenance tasks performed by students and instructors. By dynamically manipulating these lecture components, we are able to personalize the lecture videos to suit individual needs. This helps to make the videos more suitable for absorption of the subject matter by students. While full-text indexed retrieval systems have been proposed earlier, our method is more efficient as it uses all forms of media to segment and index the video. It also allows us to perform efficient contextual search with a minimum of human supervision. Future integration work will focus on increasing the automation of media analysis for the courseware and the automatic construction of media models. Additional work on an enhanced learning environment involves the creation of query mediation tools that bridge the gap between the semantically labeled media and a domain ontology. The combination of media models and query mediation will enhance insight generation and provide support for the personalized construction of knowledge by the individual student.

REFERENCES

Akutsu, A., Tonomura, Y., Ohba, Y., & Hashimoto, H. (1992). Video indexing using motion vectors. In *Proceedings of SPIE Visual Communications and Image Processing* (pp. 343-350), Boston.

Altman, E., & Wyse, L., (2005). Emergent semantics from media blending. In U. Srinivasan & S. Nepal (Eds.), *Managing multimedia semantics* (pp. 363-390). Hershey, PA: Idea Group Inc.

Bibiloni, A., & Galli, R. (1996). Content-based retrieval video system for educational purposes. In *Proceedings of the Eurographics Workshop on Multimedia, "Multimedia on the Net" (EGMM '96)*. Retrieved May 20, 2003, from http://citeseer.nj.nec.com/53258.html

Cha, G. H., & Chung, C. W. (2001). Content-based lecture access for distance learning. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2001)* (pp. 160-163).

Chambers, J. A., & Specher, J. W. (1980). Computer-assisted instruction: Current trends and critical issues. *Communications of the ACM, 23*, 332-342.

Gonzalez, R. C., & Woods, R. E. (1992). Digital image processing. Reading, MA: Addison-Wesley.

Gudivada, V. N., & Raghavan, V. V. (1995). Content-based image retrieval systems. *IEEE Computer, 28*(9), 18-22.

Hwang, J.-N., Youn, J., Deshpande, S., & Sun, M. T. (1997, October). Video browsing for course-on-demand in distance learning. In *Proceedings of the IEEE Computer Society International Conference on Image Processing (ICIP)* (pp. 530-533).

Idris, F., & Panchanathan, S. (1997). Review of image and video indexing techniques. *Journal of Visual Communication and Image Representation, 8*(2), 146-166.

Ip, H. H. S., & Chan, S. L. (1998). Automatic segmentation and index construction for lecture video. *Journal of Educational Multimedia and Hypermedia, 7*(1), 91-104.

Michalopoulos, D. A. (1976, February). A video disc-oriented educational system. In *Proceedings of the ACM SIGCSE/SIGCUE Technical Symposium on Computer Science and Education* (pp. 389-392).

Mittal, A., & Cheong, L.-F. (2003). Framework for synthesizing semantic-level indices. *Journal of Multimedia Tools Applications, 20*(2), 135-158.

Mittal, A., Dixit, S., Maheshwari, L. K., & Sung, W. K. (2003, July). Enhanced understanding and retrieval of e-learning documents through relational and conceptual graphs. In *Proceedings*

of the AIED'03 Workshop on Technologies for Electronic Documents for Supporting Learning, Sydney, Australia (p. 9).

Mittal, A., & Altman, E. (2003, January). Contextual information extraction for video data. In *Proceedings of the 9th International Conference on Multimedia Modelin*g, Taipei, Taiwan (pp. 209-223).

Wang, Y., Liu, Z., & Huang, J. C. (2000, November). Multimedia content analysis using both audio and visual cues. *IEEE Signal Processing Magazine*, *17*(6), 12-36.

Witbrock, M. J., & Hauptmann, A. G. (1997, July). Using words and phonetic strings for efficient information retrieval from imperfectly transcribed spoken documents. In *Proceedings of the 2nd ACM International Conference on Digital Libraries*, Philadelphia (pp. 23-26).

Woudstra, A., Velthausz, D. D., de Poot, H. G. J., Hadidy, F., Jonker, W., Houtsma, M. A. W. et al. (1998). Modeling and retrieving audiovisual information: A soccer video retrieval system. In *Proceedings of the Advances in Multimedia Information Systems 4th International Workshop (MIS'98)*, Istanbul, Turkey (pp. 161-173).

ENDNOTE

1 Retrieved from http://web.mit.edu/sma/ (last accessed September 4, 2004).

This work was previously published in Flexible Learning in an Information Society, edited by B. H. Khan, pp. 164-177, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 5.10
Using Multimedia and Virtual Reality for Web-Based Collaborative Learning on Multiple Platforms

Gavin McArdle
University College Dublin, Ireland

Teresa Monahan
University College Dublin, Ireland

Michela Bertolotto
University College Dublin, Ireland

ABSTRACT

Since the advent of the Internet, educators have realised its potential as a medium for teaching. The term e-learning has been introduced to describe this Internet-based education. Although e-learning applications are popular, much research is now underway to improve the features they provide. For example, the addition of synchronous communication methods and multimedia is being studied. With the introduction of wireless networks, mobile devices are also being investigated as a medium to present learning content. Currently, the use of 3-dimensional (3D) graphics is being explored for creating virtual learning environments online. Virtual reality (VR) is already being used in multiple disciplines for teaching various tasks. This chapter focuses on describing some VR systems, and also discusses the current state of e-learning on mobile devices. We also present the VR learning environment that we have developed, incorporating many of the techniques mentioned above for both desktop and mobile devices.

INSIDE CHAPTER

E-learning has become an established medium for delivering online courses. Its popularity is mainly

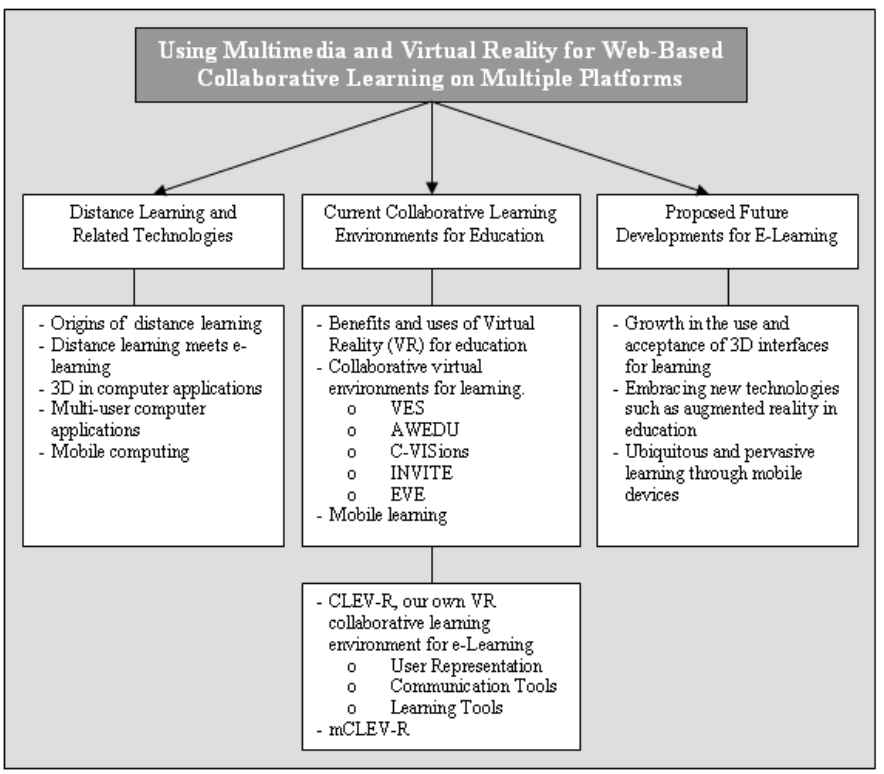
due to the convenience and flexibility it provides for users, allowing them to learn without time or location restrictions. Many different e-learning systems are currently available, the majority of which are text-based and allow users to contact the course tutor via electronic mail or discussion forums. These courses essentially offer access to a common pool of resources that allow users to gain knowledge and often qualifications. Researchers are now exploring new ways of making the online learning experience more engaging and motivating for students. Multimedia and communication technologies are being added, and together with 3D graphics, are fast emerging as a means of creating an immersive online learning experience. With the advent of mobile technologies, m-learning is showing promise as an accompaniment to online courses, offering the prospect of a modern and pervasive learning environment.

This chapter discusses the benefits 3D environments offer the e-learning community. We outline how this type of system emerged and describe some currently available systems using these new technologies. In particular, we describe in detail our own virtual reality environment for online learning and the features it provides. We discuss the extension of this system to a mobile platform so that users have anytime, anywhere access to course materials. Finally, we put forward some thoughts on future technologies and discuss their possible contribution to the development of a truly ubiquitous and pervasive learning environment.

INTRODUCTION

Distance learning has gone through a number of iterations since its introduction in the 1800s. The

Figure 1. Chapter overview



notion of distance learning grew mainly out of necessity, and helped to overcome geographical, economical, and cultural barriers that prevented people from partaking in traditional classroom-based education. Over the years a number of distance learning applications have emerged to address these issues. The evolution of such systems can be clearly linked to the technological developments of the time. This chapter focuses on giving a brief overview of the changes in distance learning from its inception to today, before concentrating on the distance learning technologies currently in use. We provide details of how the latest technologies and demand from students have led to the development of 3-dimensional (3D) e-learning systems. We also look to the future, suggesting what the next generation of technology can bring to distance learning. We pay particular attention to the need for ubiquitous and pervasive means of e-learning, and in doing so describe our own system, which uses state of the art technologies to deliver learning material to students both on a desktop computer and while they are on the move.

In the background section, we describe how distance learning has evolved from a simple postal service offered by universities to a sophisticated tool that utilises the convenience of the Internet. As the discussion progresses toward the introduction of 3D graphical environments to distance learning applications, the origins of 3D graphics and their uses are also presented. Multi-user environments for distance education is a major area of research at present, and so the latter part of the section provides a synopsis of the history of multi-user computer applications. We also present a brief discussion on the current uses of mobile technologies, which are now emerging as promising tools for e-learning.

In the main section of this chapter, we describe how recent technological advancements and requirements of students and tutors have led to a new breed of computer-based learning systems utilising the latest 3D graphics and communication tools. We detail a number of such systems outlining their strengths and weaknesses, in particular the system we are developing, which attempts to address some of these weaknesses. We describe how it uses the latest technologies to deliver a collaborative multi-user 3D learning system to students at fixed computer terminals and mobile devices. Finally we hypothesise how the current use of augmented reality (AR) technologies can be adapted to form a truly ubiquitous and pervasive learning environment. A summation and discussion of the chapter is provided in the concluding section.

BACKGROUND

This section gives the reader an overview of how distance learning has evolved from its early days as correspondence courses to the modern Internet based learning solution. It also charts the progression of computer-based 3D and multi-user tools, hinting at how they can be combined to form a new type of distance learning system. This new learning paradigm is made all the more powerful when combined with the latest mobile technologies, and a short overview of this emerging technology is provided below. Figure 2 provides an overview of the literature reviewed in this chapter, while Figures 3 and 4 highlight the main points of discussion in each subsection.

A Brief History of Distance Learning

Distance learning is a form of education that has emerged out of necessity. It is not always possible for a student and instructor to be present at the same location at the same time. Distance learning is not a new concept, and has been in use since the 1800s. Today, it can take a wide variety of forms, including correspondence courses, video and radio broadcasts, and e-learning. One of the earliest forms of distance learning was a correspondence course. Traditionally, these courses were a form

Figure 2. Background

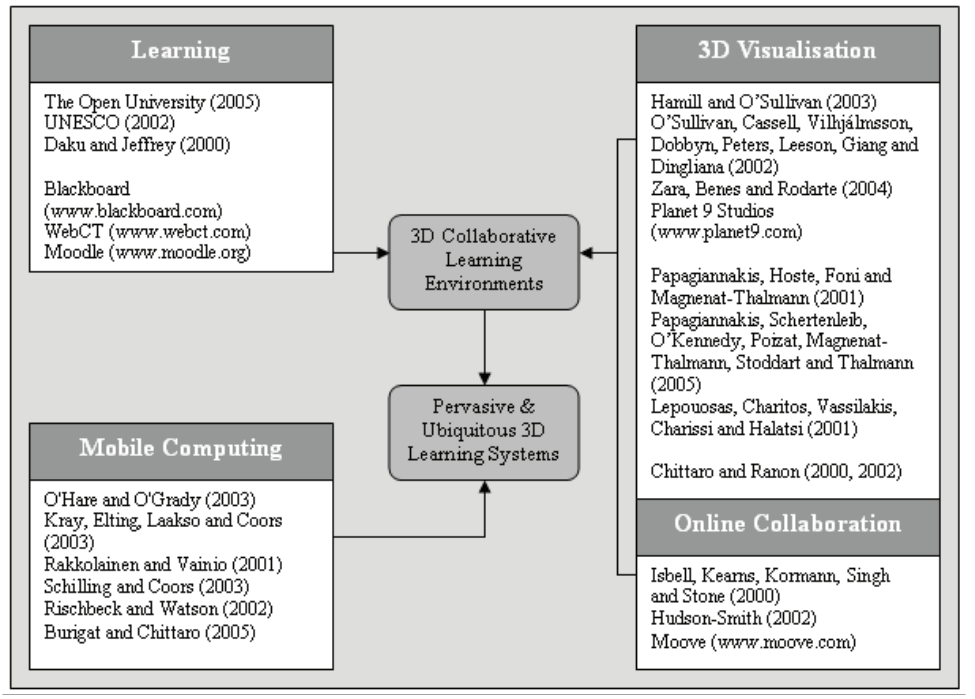
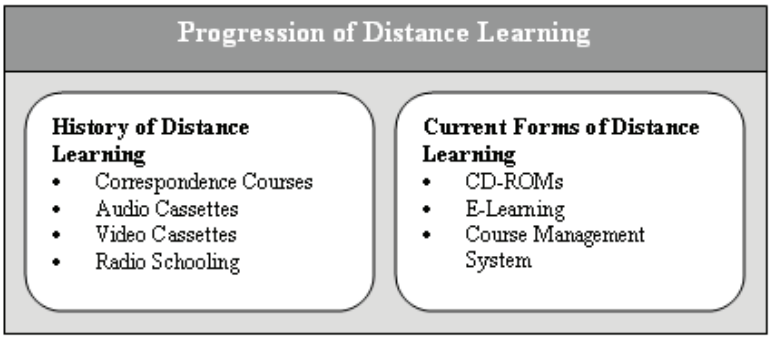


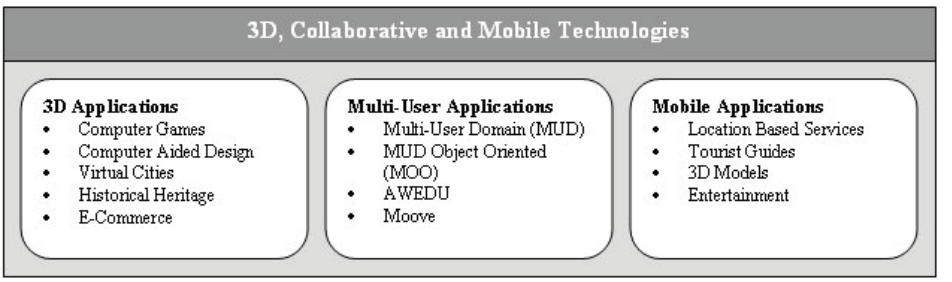
Figure 3. Progression of distance learning



of home study. Students receive printed or written course material via the postal service, complete assignments, and return them to their tutor for appraisal. These courses were later augmented with different types of media. For example, the Open University (2005) offered lectures on audio and video cassettes in the 1980s.

One major drawback of this form of distance education was the lack of interaction between student and teacher. Radio schooling provided a solution to this issue. Again, radio schooling is a form of home study, mainly used for primary and secondary education; pupils in remote locations can use a two-way radio to communicate with teachers and receive their assignments. This type of learning proved popular in Australia in the 1980s, and was particularly effective where large numbers of people lived in remote locations. Today it is being used in Africa to provide primary education for pupils in remote towns and villages (UNESCO, 2002). One of the key advantages radio schooling provided over the

Figure 4. 3D, collaborative, and mobile technologies



more traditional correspondence courses was the instant feedback between student and tutor. With the advancement of technology, this type of one-to-one distance education has now become much more attainable. Indeed, the advent and widespread growth of computer technology has introduced many additional communication and interactive features to distance learning, moving it forward into the realm of e-learning.

Distance Learning Meets E-Learning

E-learning is a term given to any form of learning which involves the use of an electronic medium to display learning material. Early forms of e-learning involved teachers demonstrating certain lessons to students through computer-based animations and simulations. CD-ROMs were then developed and soon became a popular accompaniment to textbooks. Students could use these CD-ROMs to further understand the book's content, and also as a study aid. Distance learning courses can also utilise this technology to distribute course material. The use of CD-ROMs means that the students' learning experience can be much more interactive and that learning content can be represented in different formats, including sound and video clips. For example, when studying historical events, news footage from the era can be displayed via the CD-ROM, helping to bring the learning material to life. CD- ROMs often provide interactive games (e.g., counting games) for younger students, helping them to improve their numerical skills. Automated quizzes provide instant feedback of a student's knowledge of a subject area. Daku and Jeffrey (2000) describe one such CD-ROM used for teaching, MATHLAB, a statistical application; it acts as a standalone tool for learning the functionality of MATHLAB. Experimental results have shown that students preferred learning using the CD-ROM rather than the traditional lecture/assignment format.

Over the last 10 years the popularity of the Internet as an information source has grown extensively. Its shear diffusion and convenience is ideal to disperse learning content for distance education. In general, Web sites are designed where tutors upload course material, including text, quizzes, and links to external knowledge sources. Registered students can then access this learning material and study it at their own pace. Course work can be submitted to the tutor using e-mail, and likewise students experiencing problems may contact tutors via e-mail. There are numerous examples of this type of distance learning system in use today. For example, the University of Colorado and the University of Illinois provide these kinds of courses. Indeed, many schools and universities now also use Web-based learning as an accompaniment to traditional classroom and lecture-based lessons.

Realising the importance and benefits of using the Internet for distance learning, much research is underway to improve the services and facilities

that such learning portals can offer. Initially these Web sites were a mere bank of knowledge, simply providing the course material in HTML (Hyper Text Mark-up Language) format for students to access, read, and learn. Today, they are far more sophisticated. A number of successful companies have emerged which offer online course management tools for tutors to intuitively present course notes, lecture slides, and additional material online. All management, such as access rights and course registration, are provided by these applications. Blackboard (*www.blackboard.com*) is one such course management system, designed to facilitate teachers with the management of their online courses. It has been adopted as the e-learning platform by more than 2200 institutions, including Harvard University, the University of Maine, and the University College Dublin (Wikipedia, 2005). Web Course Tools (WebCT, *www.webct.com*) is another company involved in the provision of e-learning systems. Founded in 1995, it is currently the world's leading provider of e-learning systems, with institutes in over 70 countries relying on them for e-learning software. Like Blackboard, WebCT is an authoring environment for teachers to create online training courses. While these systems tend to be extremely costly, free systems with open source code have also been developed. Moodle (*www.moodle.org*) is one such learning system, which is in widespread use at institutions such as Dublin City University, University of Glasgow, and Alaska Pacific University, and looks like becoming the industry standard. Development of Moodle has been ongoing since 1999, and already it has a large user base. It offers a range of software modules that enable tutors to create online courses. One area that Moodle tries to address is the need for pedagogical support; this aspect is largely neglected in commercially available applications. In particular it promotes the notion of constructionist learning, where a student learns from his or her own experiences, resulting in a student-centred learning environment.

The current state of e-learning, and in particular distance learning, has been outlined above. Following the success of computers as a learning tool, much research is underway to enhance a user's learning environment**.** In particular, one area that is being researched is the use of 3D graphics in these systems. It is the examination of the use of 3D graphics and multimedia, combined with collaborative tools and mobile technologies within e-learning, that forms the focus of this chapter. The remainder of this section provides a brief description of the use of 3D graphics on computers, from their early days in computer games to its current role in visualisation. We also examine the emergence of 3D graphics on mobile devices such as personal digital assistants (PDAs). This section gives an insight into the previous use of collaboration tools and multi-user interaction on computers before the next section gives a detailed review of learning systems utilising these technologies.

3D, Collaborative, and Mobile Technologies

3D Computer Applications

For some time, 3D graphics have been an established means of entertainment, in particular within computer games. In 1993, ID Software released Doom, a first person shooter computer game. Like its predecessor Wolfenstein, 3D released in 1992, Doom was built using pseudo-3D, where images placed on a 2D plane give the impression of a 3D environment. The immersive game environment included advanced features such as stereo sound and multilevel environments with increased interaction for players. The popularity of these games and the 3D paradigm led to a succession of immersive 3D computer games using this formula, and ultimately led to the worldwide acceptance of this form of immersive environment. As technology improved, the complexity of the graphics used in such computers games increased.

Today in games such as Half-Life, developed by Value Software, users take on roles of characters in stories with life-like scenes. Add-ons enable multi-user support for players to interact with each other in the game, although they may be geographically distant.

Another domain where 3D has been used for a long time is modelling; in particular engineering and architectural models can be effectively and efficiently modelled on a computer using 3D graphics. AutoCAD, developed by AutoDesk, is a computer aided drafting software application first developed in 1982 for use by mechanical engineers. It allows both 2D and 3D representation of objects to be rendered and has fast become the industry standard. This form of modelling objects enables designers and customers to see how objects will look before they are built. Today property developers often produce 3D Virtual Reality (VR) models of houses in order to entice prospective buyers. This can be taken one step further as developers produce models of how new builds will affect the aesthetics of an area. Another area of interest in the research arena is that of modelling cities. Hamill and O'Sullivan (2003) describe their efforts in producing a large-scale simulation of Dublin City. The goal of their research is to allow users to navigate freely around the streets of Dublin. The city can be used as a test-bed for related work, such as simulating crowds in virtual environments (O'Sullivan, Cassell, Vilhjálmsson, Dobbyn, Peters, Leeson, et al, 2002). Zara, Benes, and Rodarte (2004) have produced a model of the old Mexican city of Campeche. This Internet-based application acts as a tourist aid and promotes the area as a tourist attraction. The area of urban modelling has much commercial interest. Planet 9 Studios (*www.planet9.com*), based in San Francisco and set up in the early 1990s, produces accurate and intricate 3D urban models for use in such diverse activities as homeland defence, military training, and tourism. It is tourism that has long been a driving force behind much of the development of the VR cities and has also led to a number of similar applications aimed at visitors and tourists. MIRALabs in Geneva have been investigating the area of virtual heritage. Much of this research focuses on enabling realistic models of historical sites to be created efficiently and in a means suitable for use on basic machines (Papagiannakis, Hoste, Foni, & Magnenat-Thalmann, 2001). Today, the work of MIRALab in the field of historical reconstructions involves the use of mixed realities; that is, augmenting real-world sites with VR 3D models of people and artefacts that would have once been there (Papagiannakis, Schertenleib, O'Kennedy, Poizat, Magnenat-Thalmann, Stoddart, et al, 2005). The benefits of using VR to host exhibitions in museums are outlined by Lepoursas, Charitos, Vassilakis, Charissi, and Halatsi (2001), where details on designing, such an exhibition, are also provided.

Several of the projects mentioned above use special features to allow users to interact with the environment. Often the environment acts as a 3D-Graphical User Interface (GUI) to access underlying information. E-commerce or purchasing products and services using Web sites on the Internet have long been popular. The number of people buying through this medium has increased dramatically over the last five years. According to Johnson, Delhagen, and Yuen (2003), online retail in the United States alone will have reached $229.9 billion by 2008, and so this is a natural area for researchers to investigate. While research has contributed to many improvements in this form of shopping, one area that is still emerging is the use of a 3D interface to browse for goods. Chittaro and Ranon (2002) have designed a virtual environment mimicking a traditional supermarket, where virtual products are placed on virtual shelves. Users navigate the store, selecting items they wish to purchase. They argue that this is a more natural way for consumers to shop online because it is more familiar to shoppers compared to lists of available items. These 3D department stores can be tailored and personalised to an individual user (Chittaro & Ranon, 2000). This

can involve personalising the look and feel of the environment and user profiling to position relevant items in prominent positions. The extension of this 3D store paradigm to a multi-user platform is also discussed and proposed, where a number of shoppers are present in the one store. This, however, has drawbacks and provides difficultly in adapting the store for individuals. The next section provides a synopsis of the origins of multi-user computer applications and provides some details of their current uses.

Multi-User Applications

One of the earliest forms of multi-user interactivity using computers took the form of multi-user domains, known as MUDs. MUDs first appeared in 1978, and being primarily used for computer gaming purposes, quickly became popular. In a MUD, a text-based description of an environment is provided, with each user taking on the role of a character in that environment. Interaction with the system and with other users is achieved by typing commands in a natural language. Usually a fantasy game featuring goblins and other creatures, the purpose was to navigate through the virtual environment, killing as many demons as possible. While initially developed for entertainment, people saw the possibility of using this technology for other purposes, notably as a form of distance learning and as a means of virtual conferences. This move away from the use of MUDs for gaming led to the developments of a MOO (MUD Object Orientated) at the University of Waterloo in 1990. Again, MOO systems are text-based virtual environments, which were initially an academic form of a MUD. Isbell, Kearns, Kormann, Singh, and Stone (2000) discuss LambdaMoo, one of the longest running MOOs, created by Pavel Curtis in 1990. Today, it offers a social setting for connected users to engage in social interaction with similar functionality to that found in a chat room or online communities.

Improvements in technologies, along with increases in Internet connection speeds, have enabled a move away from the traditional text-based environments to more graphical-based communities. These communities, such as Active Worlds (Hudson-Smith, 2002), offer the same interaction as chat rooms. However, the rooms are designed as physical spaces complete with scenery and furniture. Each user is shown on-screen in the form of an avatar, a graphical representation visible to other users. A further extension of this type of environment is seen in the 3D online world of Moove (*www.moove.com*). Here, users maintain their own rooms in the 3D environment, which they can decorate to their own tastes and use to host chat sessions. Voice and video chat can also be used in many of the modern online 3D communities. In recent times, this type of 3D environment has been used for education, and this will be discussed later in this chapter.

Mobile Applications

In recent years, the use of mobile devices such as PDAs and cell phones has become prevalent, and this has led to a lot of research into providing applications on these devices. The widespread introduction of wireless networks has increased the use of these devices dramatically, and has helped fuel research in this area. People can now browse the Internet, send electronic mails (e-mails) and access their personal files while on the move. Many different applications have been developed for this mobile platform, the most popular of which utilise a global positioning system (GPS). These applications use a system of satellites and receiving devices to compute positions on the Earth, and therefore enable people to gain information relative to their position in the world. They are thus used mainly in providing location-based information to users, and in particular many tourist applications have been developed to help people find their way around foreign cities. For example, O'Hare and O'Grady

(2003) have developed a context-aware tourist guide, which tracks a user's position and displays multimedia presentations for different attractions as the user approaches them. Their system uses images, videos, sound, and text to give the tourist information about the attraction. Kray, Elting, Laakso, and Coors (2003) developed an application to provide boat tourists with route instructions and information on services nearby.

A recent development in mobile technologies is the use of 3D graphics on these devices. Many large-scale 3D models have been successfully developed for use on laptop computers, but the challenge now is to extend these to smaller platforms, such as PDAs and even mobile phones. Rakkolainen and Vainio (2001) developed a Web-based 3D city model of Tampere and connected it to a database of relational information so that users can query the system about services available in the city. They customised it for mobile users by integrating a GPS receiver and have developed a fully working version for laptops. However, their 3D model is much too large to run on smaller devices such as PDAs, so they use only images and Web pages for these devices. Schilling and Coors (2003) have developed a system that provides a 3D map to help present route instructions to mobile users. In their model, landmarks and buildings of importance are visualised in detail through the use of textures, while less important buildings are rendered in grey. A user trial was carried out on a laptop running a mobile phone emulator. Results proved positive, with most users stating they would prefer to use the 3D maps than 2D paper maps and guidebooks. However, users did suggest that the 3D model should be more detailed and more realistic.

The major problem for extending these models to smaller mobile devices is their size in relation to processing power available, together with the users' desires for more detail and realism. Many researchers are exploring ways to achieve this using various culling techniques, parallelism, and information filtering (Burigat & Chittaro, 2005; Rischbeck & Watson, 2002). Smaller 3D models can, however, be displayed on smaller devices and are used in a variety of games on handheld devices and mobile phones, providing entertainment for their user. ParallelGraphics, a world leader in the provision of Web3D graphics solutions, describe uses of 3D for mobile devices, such as sales and marketing, real estate, and field maintenance.

The technologies presented in this section have been widely accepted by developers and computer users. As the next section shows, they have recently received attention from the educational research community. They show promise as a means of presenting learning material to students in an engaging and motivating way. A number of such systems using these technologies are presented in the following section before we discuss our own learning system, CLEV-R, which provides solutions to some of the issues that existing learning systems fail to address. In particular, the need for a range of communication and collaborative tools for both learning and socialising is dealt with. Also presented is a pervasive ubiquitous learning environment we developed for use on mobile devices as an accompaniment to the desktop system.

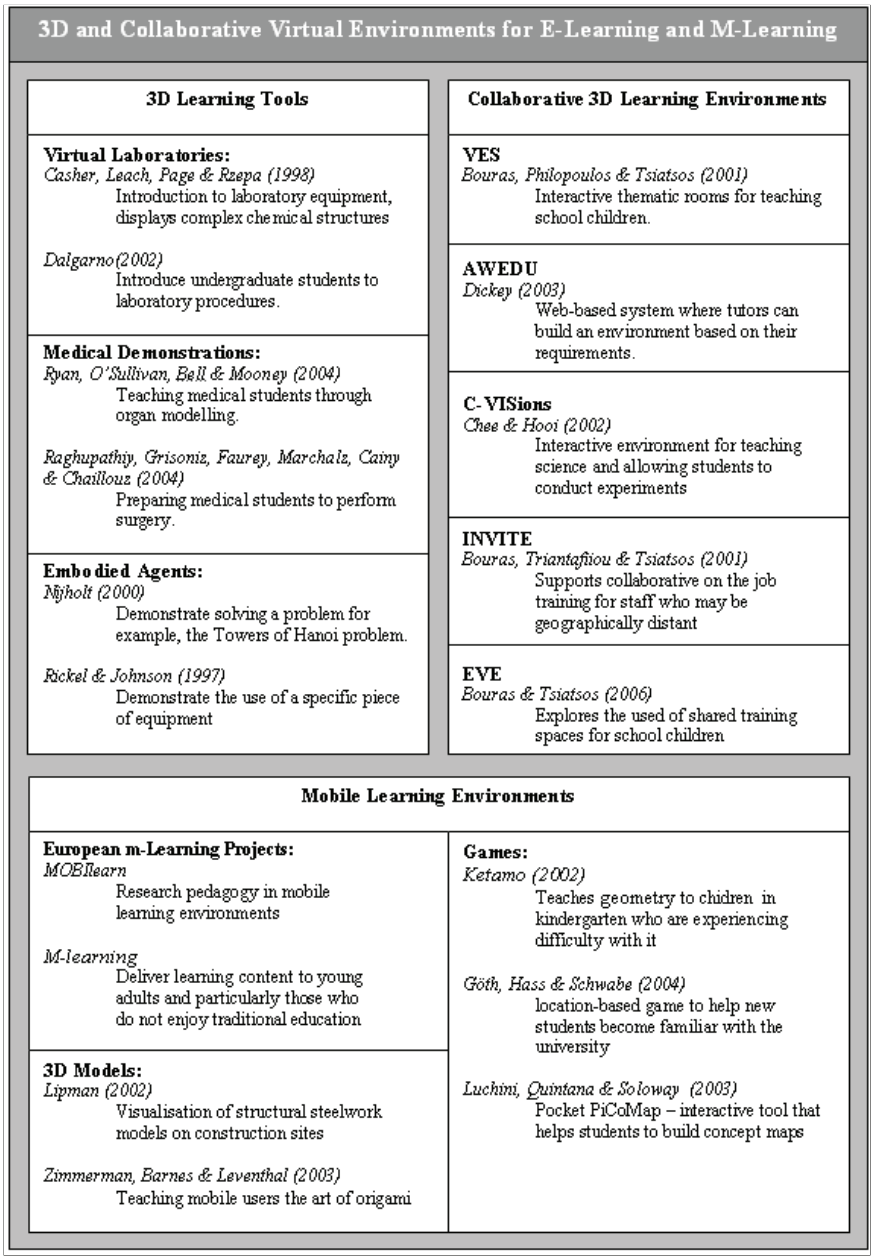
3D AND COLLABORATIVE VIRTUAL ENVIRONMENTS FOR E-LEARNING AND M-LEARNING

This section details the use of 3D as a learning aid. Firstly, we consider its use as a visualisation tool, and then discuss how multi-user technologies are being combined with 3D graphics to create effective online learning environments. We also discuss some current research into the provision of learning tools on mobile devices. Figure 5 provides an overview of the topics presented below.

3D Learning Tools

In addition to the uses of virtual reality (VR) and 3D graphics discussed in the previous section, these techniques have also been extended in vari-

Figure 5. 3D and collaborative virtual environments for e-learning and m-learning



ous ways for use in education. The ability of these tools to model real-world objects and visualise complex data makes them an ideal learning tool. Users can explore these objects, interact with them, and discover their various features. Furthermore, the visualisation of complex data can greatly aid a person's comprehension of it. Thus, these models provide users with an intuitive way to learn about natural objects by presenting them in a visually appealing way. As such, many 3D resources for education have been developed, both for online and individual applications.

Scientific and engineering visualisations use VR to represent complex chemical structures

and to present experimental data in a more visual manner in order to gain a better understanding of the results. Casher, Leach, Page, and Rzepa (1998) describe the use of VR for chemical modelling, and outline the advantages that animation can bring to these models. They also describe how a virtual laboratory can introduce students to various laboratory instruments. In addition, Dalgarno (2002) has developed a virtual chemistry laboratory that allows undergraduate students to become familiar with the layout of the labs, and also to learn about procedures to follow while in the laboratory. VR has more recently been introduced in the field of medical training. Its use varies from modelling different organs and allowing students to interact with them freely to developing training application for specific procedures. Examples include Raghupathiy, Grisoniz, Faurey, Marchalz, Caniy, and Chaillouz (2004), who developed a training application for the removal of colon cancer, and Ryan, O'Sullivan, Bell, and Mooney (2004), who explore the use of VR for teaching electrocardiography. The major advantage of using VR in medicine is that students can repeatedly explore the structures of interest and can interactively view and manipulate them. Real training cases can be hard to come by, and so this extra practice and experience can be invaluable. Also, patients are not put at risk by having inexperienced students carry out procedures on them.

3D models can also be particularly useful in teaching younger students. Many games have been developed using 3D images that the user must interact with in order to learn a certain lesson. Interactive models increase a user's interest and make learning more fun. 3D animations can be used to teach students different procedures and mechanisms for carrying out specific tasks. Some researchers have combined the benefits of 3D and software agent technologies to provide intelligent models to teach certain tasks. For example, Jacob is an intelligent agent in a VR environment that guides the user through the steps involved in solving the towers of Hanoi problem, as described by Nijholt (2000). By following the directions of Jacob, the users learn how to solve the problem themselves. Likewise, STEVE, described by Rickel and Johnson (1997, 1999), is an intelligent agent that has been developed for use in naval training to show individuals or groups of students how to operate and maintain complex equipment. STEVE can demonstrate certain tasks and then observe while users carry out these tasks, correcting them when mistakes are made.

Collaborative 3D Learning Environments

VR also allows for the development of complete virtual environments that users can "enter" and navigate through as if it was a real environment. The most immersive of these environments require the user to wear a head mounted display and tracking gloves, while other VR environments are displayed on desktop computers, where users interact through the mouse and keyboard. As shown in the previous section, virtual environments like these have evolved from computer games, but are fast emerging in other areas such as e-commerce, chat-rooms, and indeed education. E-learning in particular is an ideal target for the development of an immersive VR environment. Here an entire VR environment is designed where all the learning takes place. This kind of system represents a shift in e-learning from the conventional text-based online learning environment to a more immersive and intuitive one. Since VR is a computer simulation of a natural environment, interaction with a 3D environment is much more intuitive than browsing through 2D Web pages looking for information. These environments tend to be multi-user, exploiting the notion of collaborative learning where students learn together. The benefits of collaborative learning have been researched extensively and are outlined in Laister and Kober (2002) and Redfern and Naughton (2002). The main advantage of this type of learning is that users no longer feel alone or isolated. This

feeling of isolation can be particularly prevalent in online learning, where students do not attend actual classes or lectures. Thus, multi-user learning environments have proven very popular for online learning. The VES, AWEDU, C-VISions, EVE, and INVITE systems all concentrate on developing collaborative learning environments using VR to further immerse students in their learning. The following paragraphs outline the main features of these systems before we discuss our own research and what it has to offer.

In 1998, Bouras, Fotakis, Kapoulas, Koubek, Mayer, and Rehatscheck (1999) began research on virtual European schools (VES), the goal of which was to introduce computers to secondary school students and encourage teachers to use computers in the classroom. VES uses 3D to provide a desktop immersive environment. A different room in the 3D environment is used for each school subject, and these themed rooms provide information about the specific subject in the form of slide shows and animations, as well as links to external sources of information. The VES project was carried out in conjunction with book publishers, and these publishing houses provided much of the content that is displayed in the environment. VES is an example of a multi-user distributed virtual environment (mDVE). In an mDVE, more than one person can access the environment at the same time, and users are aware of one another. In the VES environment users can "talk" to each other using text chat facilities. The evaluation of VES took the form of questionnaires, which students and teachers completed. The results, which are presented in (Bouras, Philopoulos, & Tsiatsos, 2001) show that navigation in the 3D environment was difficult, the user interface was old fashioned, and there was not enough content to keep students amused and entertained. These points were taken on board and the system was improved. When launched, VES was used in 4 countries and had the cooperation of more that 20 publishers.

The Active Worlds Universe, a very popular and powerful Web-based VR experience, is a community of thousands of users that chat and build 3D VR environments in a vast virtual space. As discussed earlier, it provides thousands of unique worlds for shopping, chatting, and playing games. In 1999, an educational community known as the Active Worlds Educational Universe (AWEDU) was developed (Dickey, 2003). This is a unique educational community that makes the Active Worlds technology available to educational institutions. Through this community, educators can build their own educational environment using a library of customisable objects, and can then place relevant learning material in their environment. Through these environments, users are able to explore new concepts and learning theories and can communicate using text-chat. Users are represented in the environment by avatars which help them feel better immersed in the educational environment. Students from all over the world can be connected through this system, and it therefore aids cultural sharing and social learning. The AWEDU environment is extremely versatile and may be used for a number of types of learning. Dickey (2003) presents the use of the environment as a form of distance education within the university. Riedl, Barrett, Rowe, Smith and Vinson, (2000) provide a description of a class held within the Active Worlds environment. The course was designed for training teachers on the integration of technology into the classroom. Nine students took part in the class; their actions and group discussions were recorded during their online sessions and, along with results from questionnaires, are presented in the paper. While the majority of students were pleased with the freedom the virtual environment offered, not all adapted to this new form of learning. The evaluation discovered that one of the major benefits of this type of learning was that students were aware of the presence of others in the shared environment and this interaction with others kept students interested and motivated.

C-VISions was launched in 2000, and is a collaborative virtual learning environment that concentrates on supporting science learning. The system presents learning stimuli that help school children understand fundamental concepts from chemistry, biology and physics. C-VISions encourages active learners; students can run science experiments in the virtual world and view the outcomes as they change simulation parameters. Chee and Hooi (2002) describe their physics environment and in particular the Billiard World, a simulation to help students learn about mass, velocity, acceleration, conservation of momentum, friction, and the coefficient of restitution. This world contains a billiard table with two balls and a cue stick. Users can interact with a number of "live" objects that are provided within the world. For example, the cue stick can be aimed at a ball and then used to strike it. Students can replay the most recent simulation and can view the plotting of graphs of that event synchronously. This helps the students see the relation between their action and how it is plotted on a graph. It therefore helps them to understand the graph representations. Users can navigate around the world and change their viewpoints using buttons provided. The system is multi-user, and so events happening in one user's environment are propagated to all other connected users. Users may share video resources, and shared electronic whiteboards are also provided. The system provides a Social World where students can mingle and student-student communication is supported through text and audio chat. Chee (2001) describes a preliminary evaluation of the first prototype of this system. The study revealed that all students found the system "an enjoyable way to learn" and each felt they gained a better sense of understanding about the subject matter. Some problems using the collaboration tools were highlighted. For example, students found it difficult to work together on group tasks. This was put down to inexperience using the tools. While this study was small, involving only three students, the results proved encouraging that this type of 3D environment has something to offer students.

In April 2000, a consortium of companies and institutions began research into the design and development of a multi-user 3D collaborative environment for training. The goal of the environment was to support group learning, in particular on-the-job training, without the need for all those involved to be in the same location at the same time. Moving away from the traditional videoconferences, this system known as INVITE, the Intelligent Distributed Virtual Training Environment, had a fundamental objective of making people feel that they are working as a group rather than alone in front of a computer. The technologies required for such a system are described by Bouras, Triantafillou, and Tsiatsos (2001), along with implementation issues of the multi-user architecture. The project focuses on the importance of a social presence and the general sense of belonging within a learning environment, presenting the notion of photo-realistic avatars as a way to achieve this. The system design allows synchronous viewing of e-learning content within the 3D environment through a presentation table. Users can see pictures, presentations, 3D objects and prerecorded videos simultaneously, and collaboration is provided through application sharing. An initial prototype of the system was developed, and a first evaluation showed that INVITE could be a powerful tool for collaborative learning with test-users enjoying learning in the virtual environment. The INVITE project terminated prematurely, and so the main contribution it made to the area of virtual learning environments was a detailed system specification and outline of features that should be included in such a system.

The research group from the University of Patras in Greece who were involved in the development on the INVITE Project continued their work, leading to the development of EVE (Educational Virtual Environments). Like INVITE, EVE is a Web-based, multi-user envi-

ronment that explores the use of shared virtual spaces for training. Their system addresses two main challenges. The first was a technological challenge to develop a learning environment that resembles the real world and that provides additional functionality to enhance the users' experience. Secondly, the pedagogical challenge of making an educational model that contributes in the most efficient way to the distribution of knowledge. EVE is organized into two types of place for each user, their personal desk space and the training area. The personal desk refers to a 2D place where all the asynchronous features of the system relating to that user can be accessed. Thus a user can access course and user information, upload and download files, view and reply to personal messages, and manage their profile. The training area is the virtual classroom where learning takes place, and consists of a presentation table, a whiteboard, and avatar representations for all connected students and a tutor. Features such as application sharing, brainstorming, and text and audio communication are also supported. The tutor has control over the course, learning material, and students. They decide what course content is displayed on the presentation board and when students may ask questions, and can also assign students to breakout session rooms during an e-learning class and monitor their text chat sessions. An evaluation of the system, provided by Bouras and Tsiatsos (2006), shows that test-users found the system interesting and promising for e-learning. The test users were chosen from a number of Greek schools and a teacher from each selected school also evaluated the system. The users' social presence and the intuitive virtual environment were highlighted as advantages of the system. Collaboration tools such as audio and text communication, application sharing, and visualisation on the presentation table also proved popular. Overall the feedback was positive, with both students and teachers seeing the appeal and usefulness of the 3D paradigm. Feedback was also taken from students about possible improvements that could be made to the system. The introduction of facial expressions, along with tool tips for assisting during navigation, are discussed. The future work of the EVE project therefore involves the implementation of these changes along with the addition of new rooms to the environment to support smaller groups for project work.

Mobile Learning Environments

The use of mobile devices for learning, termed m-learning, has been another area of interest for researchers of late. Their portable nature makes them convenient for many people to use while on the move. Therefore, the extension of e-learning to these devices seems a natural progression. While laptop computers are widely popular and capable of delivering large amounts of information efficiently, smaller mobile devices, such as PDAs, also show promise for learning. Oliver and Wright (2003) outline the main advantages of PDAs as their light weight and portability, their ease of use, and their low cost. Also, most are wireless enabled. Csete, Wong, and Vogel (2004) accredit the functionality provided on these devices as a reason for their growing popularity. Most mobile devices now include an address book, calendar, to-do list, and memo pad. Wireless enabled devices provide e-mail and Web browsing, and most support flash, audio, and movie files. Indeed, their functionality is continually increasing as companies are now developing versions of their software for these devices. Disadvantages of these devices, such as small screen size and limited processor power and memory, mean that applications for them must be lightweight and content needs to be adapted for this new platform. These drawbacks are, however, outweighed by their inexpensive and convenient nature, which makes them the ideal target for a learning application. The major advantage this mobile platform brings to e-learning is that students have "anytime-anywhere" access to course material.

Much research is now being carried out into providing services on these mobile devices for learning. The MOBIlearn project is a worldwide European-led research and development project exploring learning through a mobile environment (*www.mobilearn.org*). This project concentrates on creating pedagogy for learning in these environments, and looks at the adaptation of existing e-learning content for mobile devices. Their main objective is to enable content delivery for adult learning and professional development through collaborative spaces, context awareness, and adaptive human interfaces. M-learning is another European research and development programme that aims to deliver learning content to young adults who are no longer taking part in education or training (*www.m-learning.org*). In particular, they target those who are unemployed or homeless and those who do not enjoy traditional education. To engage the user in learning, themes of interest to young adults are used and are presented in the form of interactive quizzes and games. Modules include activities designed to develop aspects of literacy and numeracy. They have developed a number of learning tools ranging from interactive quizzes for teaching languages and driver theory test to giving the learners access to online Web page building and community tools.

Many researchers see games and interactive challenges as the way forward into mobile learning. Ketamo (2002) has designed a game for handheld devices that teaches geometry to 6-year-old kindergarten kids. The system proved effective and in particular helped low-skilled students understand geometry better. Göth, Hass, and Schwabe (2004) have developed a location-based game to help new university students become familiar with the university and its surroundings. Students are grouped into teams and have to carry out a number of tasks at certain locations on the campus. Pocket PiCoMap, as described in Luchini, Quintana, and Soloway (2003), is another interactive tool for mobile learning which helps students build concept maps (i.e., graphical representations of complex ideas and the relationships between them). Students draw a link between two concepts on their map and can then add a descriptive label to this linking edge. An English sentence describing the visual representation is dynamically created and displayed, thus helping the students understand how the relationship between the concepts is interpreted.

Some applications using 3D graphics on mobile devices are now also being developed offering a wide range of services to their users. For example, Lipman (2002) explored the use of 3D models on mobile devices for the visualisation of structural steelwork models on construction sites. Zimmerman, Barnes, and Leventhal (2003) designed an effective system for teaching mobile users the art of origami. A 3D model is provided showing the different steps involved in creating a particular shape. The user follows a set of instructions and uses the 3D model as a visual aid to gain a better understanding of the action they are to perform. This rendering of 3D graphics on mobile devices is an area of interest to us, and together with the potential of these devices to provide tools for collaboration, we feel a mobile learning system with these technologies could be very effective.

Above, we have outlined ways in which 3D graphics have been used for learning. Initially, 3D was used as a means of training within specific fields. For example, it proved popular for teaching medical students to perform surgery. More recently, 3D graphics have been amalgamated with multi-user technologies to form complete learning environments. A number of projects using this technology have been identified and discussed above. Within these environments, students take on the role of a character and navigate through a virtual on-screen location to access course notes and interact with each other. Each of the systems described have their own merits and limitations. Building on the strengths of the systems above and proposing a solution to their limitations, we have developed a system that recognises the importance of social learning as

part of an individual's education. This aspect was not fully addressed by other systems. Many of the traditional text-based e-learning systems discussed previously are aimed at third level students, providing diploma and degree qualifications. To date, no collaborative VR learning environment has been developed to solely cater to this particular market. Our research investigates the benefits that a 3D learning environment can bring to this domain. As the prevalence of mobile devices increases, their use as a learning tool is now being researched. Above, we have presented some currently available m-learning systems, and we too are exploring this area. In the next section we discuss our research, as we develop both a desktop 3D collaborative learning environment and a mobile application to supplement this. In particular, our system examines the use of 3D graphics in conjunction with various collaborative tools to act as a medium for learning.

COLLABORATIVE LEARNING ENVIRONMENT WITH VIRTUAL REALITY (CLEV-R)

The system which we are developing is entitled collaborative learning environment with virtual reality (CLEV-R) and addresses problems with current e-learning systems. The main objectives of the research underlying the development of CLEV-R include:

- Exploring the use of a 3D multi-user environment for e-learning, both to supplement traditional learning and for use in distance education.
- Supporting both social interaction and collaboration among system users.
- Developing the system so that it is cost-effective and requires minimal software on the client side.
- Exploring the extension of 3D interfaces for learning to mobile devices.
- Evaluating the resulting systems in terms of usability and effectiveness as a learning solution.

The following scenario indicates some issues faced by people wishing to take part in online courses. It highlights how a system like CLEV-R can address many of the concerns which people experience when using distance learning tools online.

Sample Scenario

Mary has been working for several years as an administrator in a legal firm. She wishes to further her career by obtaining a professional qualification. As she is working full time, she is unable to attend a university. Her employer suggests that she takes an online course. She has reservations about doing this because she knows of others who have found it difficult to complete courses online. They found them to be challenging; the lack of contact with others was isolating and it was difficult to maintain motivation for the duration of the course. Her friend Rachel recommends CLEV-R, an e-learning system that she found very convenient. She completed a business course using this system and said it was an enjoyable way to learn. Rachel attended university once a month; however, the rest of the course took place in a 3D virtual environment online. CLEV-R is also available on mobile devices and so she often used her PDA to access course content and communication tools while on the move. She particularly liked the collaborative aspects of CLEV-R and used them for both learning and socialising with others. Mary is convinced that this is an ideal solution for her learning needs.

The problem that Mary faced is a typical one, encountered by many people wishing to take up online courses. The social isolation, ennui, and lack of support within e-learning applications are all issues which we are addressing through the development of an online collaborative learning

environment that uses 3D graphics and VR to engage and motivate the user. Our VR environment for e-learning concentrates on providing collaborative tools so students can work, learn, and socialise together (Monahan, McArdle, Bertolotto, & Mangina, 2005). Mimicking a real university, it consists of a central common area surrounded by lecture rooms, meeting rooms, and social rooms. Learning materials are presented within the environment through various multimedia techniques and communication controls, such as text and audio chat, allow students and tutors to converse easily (Monahan, McArdle, & Bertolotto, 2005). The following paragraphs outline some of the most important features of the system and explain their use in an e-learning application.

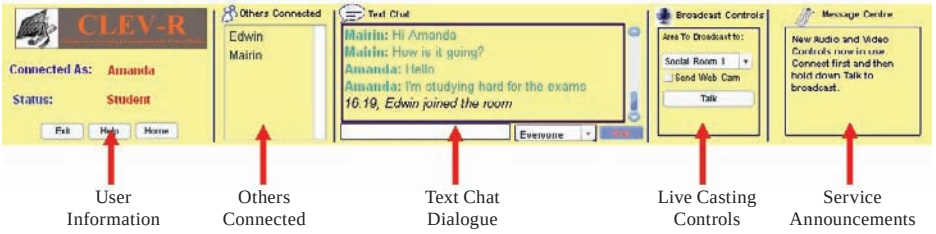
User Representation

One of the main disadvantages people see with existing e-learning applications is the lack of social presence and the feeling of isolation that can be experienced while partaking in an online course. Thus, one of the primary objectives of our environment is to remove this sense of loneliness and create a greater sense of community within each online course. The basis for our online collaborative learning environment is to facilitate multi-user support, allowing many users to be present and navigate around the same environment simultaneously. It is also vitally important that users of the environment are aware of all other connected users at any one point in time. Users are represented within the system by avatars. Upon registering for a course, each student and tutor is required to select a character to represent him or her in the VR environment. This 3D character is the user's on-screen persona for the duration of a course. In order to create an effective learning community, it is imperative that these avatars are distinctive for each individual user. In this way, avatar representations allow users of the system to recognize others, and hence feel a social presence in the learning environment. Applying different clothing or hairstyles to each character can create unique avatars.

Communication Tools

In order to support collaboration, communication technologies are imperative. In fact, it is the lack of support for real-time communication that we feel is a major drawback in current e-learning systems. In any learning scenario, it is imperative that students have a lot of communication with course tutors and also with their fellow students. Much research has been carried out to determine the importance of communication in learning, and it been has shown that students often learn from each other in informal chats as well as from lecture content (Redfern et al., 2002; Laister et al., 2002). Communicating with others who are partaking in the same course makes students feel more involved in the learning environment and removes any sense of isolation that may occur in single-user learning environments. As such, a major aspect of CLEV-R is the provision of multiple communication methods. The communication controls for CLEV-R are provided in a graphical user interface (GUI), as shown in Figure 6. Text and audio chat communication is supported. Students can send both public and private messages via text-chat and can broadcast audio streams into specific areas of the VR environment. Also users may broadcast Web-cams into the 3D environment and so have real-time face-to-face conversations with other connected users. The avatars in our system are enabled with gesture animations, which are a further form of communication. For example, avatars can raise their hand if they wish to ask a question and can also nod or shake their head to show their level of understanding of a certain topic. Of course users can also communicate asynchronously with others via e-mail.

Figure 6. The communication controls of CLEV-R



Interactive Tools

Another common problem with Web-based learning environments is that the learning content is primarily presented through various forms of text, including word files, PDF documents, HTML, and so forth. While these may be effective for presenting the learning material, they do not portray course content in a motivating or engaging way for the students. Thus, within the development of CLEV-R, we provide different multimedia methods for presenting course content within the learning environment. The system supports features such as PowerPoint slides, movies, audio, animations, and images. Rather than downloading these media files to the students' own PC, they can be experienced from within the virtual environment in real-time with other students. Many different features are available within the various virtual rooms of CLEV-R to support these file types. These are outlined in Figure 7, and the remainder of this section describes the different areas in our virtual university and the support for learning that each provides.

Lecture Room

The lecture room is the virtual space where most of the tutor-led synchronous learning occurs, and it supports a learning style similar to traditional

Figure 7. Virtual university structure within CLEV-R

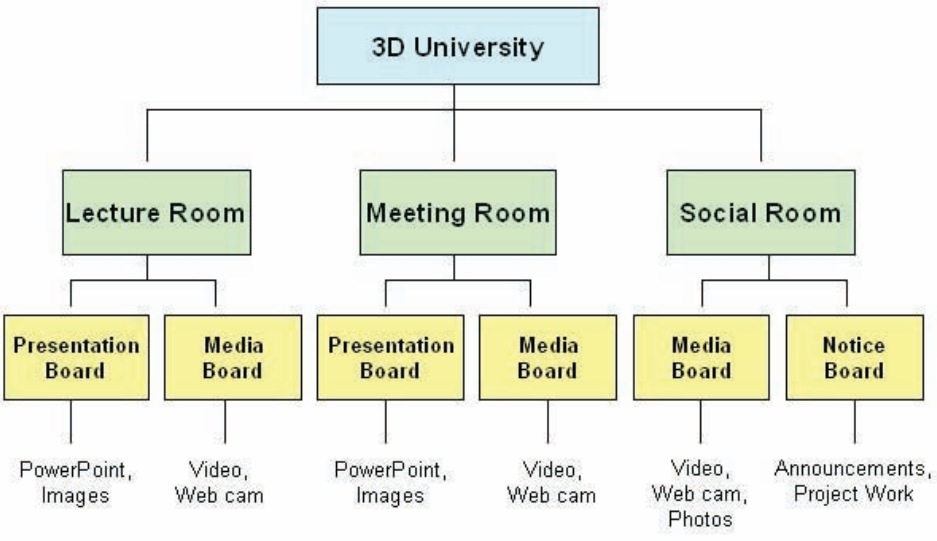
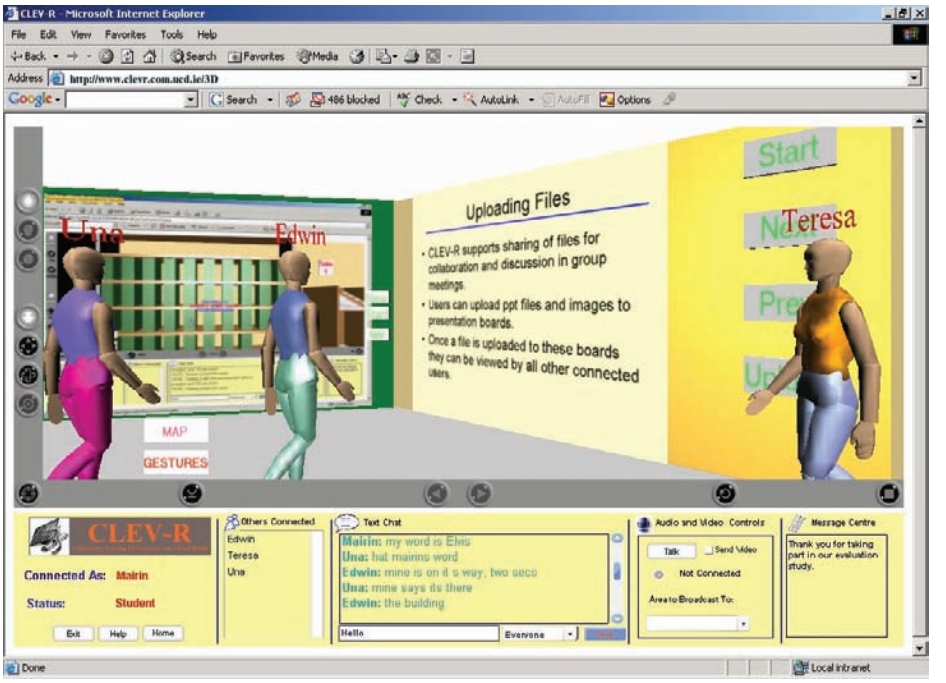


Figure 8. An online lecture taking place within CLEV-R

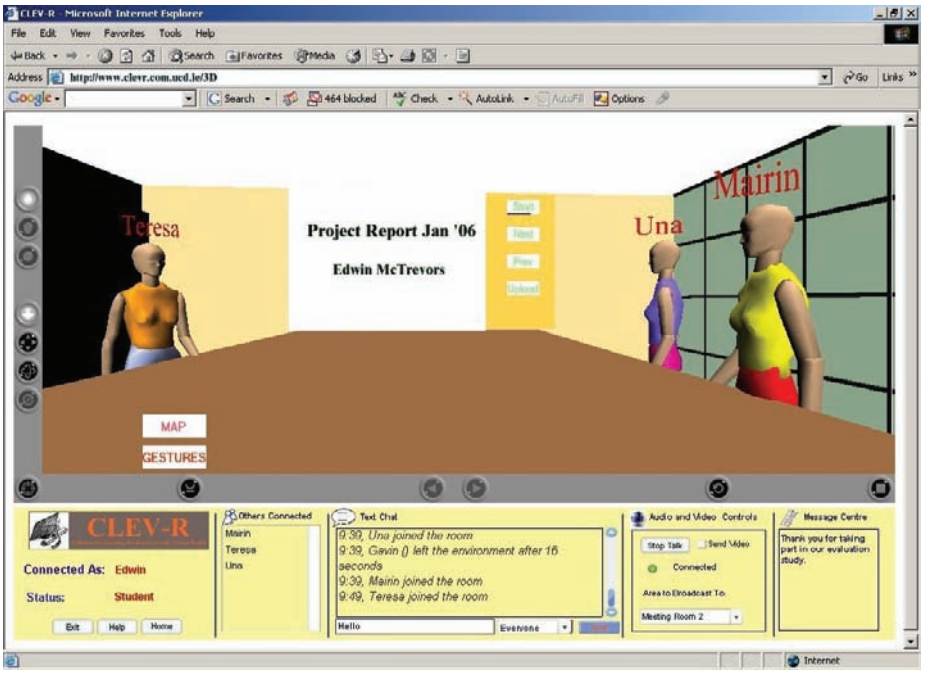


classroom-based education. This room provides several features to enable a tutor to present learning material to a number of students simultaneously. An example of an online lecture can be seen in Figure 8. A media board is provided, where the lecturer can upload both audio and video files. Where appropriate, the lecturer also has the option of streaming live Web-cam feeds into the lecture room. This can be used for demonstrating more practical aspects of a course or as a video conferencing tool for guest speakers. Lecture slides, such as PowerPoint files, can be displayed on a presentation board, which also supports images files, including GIFs, JPGs, and PNGs. The tutor controls this presentation board and can progress through slide shows displayed here. Once the tutor changes the current slide, it is changed in the worlds of all connected students. In this way, students are continually kept up-to date about the state of the environment and are always aware what learning content is currently displayed. The tutor can also use the streaming audio facility to provide live audio commentary to accompany the presentation and address any questions raised by students.

Meeting Rooms

As one of the main focuses of CLEV-R is collaboration, it is very important to provide an area for this collaboration to take place. While the entire environment can be used for collaboration, designated areas of the environment provide additional functionality for groups of students to meet and work together. The meeting room as shown in Figure 9 provides a similar set of tools found in the lecture room. Students can use audio and text messages to communicate their ideas with each other. A presentation board allows students to upload their own material to the room for discussion. Each student can bring slideshows, animations, and media clips for discussion and viewing by the entire group. Live video can also be streamed into this room via a student's Web-cam to aid with the discussion.

Figure 9. Media board displaying a video within CLEV-R



Social Rooms

Social interaction is a key component of CLEV-R, and therefore nominated areas of the 3D university have been specifically created for users to mingle and partake in informal conversation with each other. While students can use these areas to discuss the course they are attending, they can also use them for social purposes. In a similar way to the meeting rooms, small numbers of students can gather together to share their experiences and stories as well as photos, pictures, and movies. Social rooms exist where a media board is available for students to upload images, videos, and Web-cam broadcasts. A centrally located lobby also serves as an informal setting, where students can chat and build rapports with others. Here users can talk informally about the course material. Students can display their project work on special notice boards provided; others can then peruse these posters at their own pace and in their own time. Students can also place advertisements for upcoming events and other general notices in this common area.

Library

As CLEV-R is primarily a learning environment, it provides easy access to learning material through a library. The library contains a bookcase and a number of desks. Lecture notes, which have been uploaded to the lecture room by the tutor, are automatically represented in the form of a book in the bookcase. When a student clicks on one of the books, the lecture notes associated with that book are displayed on a desk in the library. The student can then peruse the notes in situ within the 3D environment or download them to their own computer. The bookcase in the library also contains a number of links to external information sources such as online dictionaries and encyclopaedias.

Evaluation and Discussion

A usability study has been conducted to obtain user feedback on the CLEV-R system and also to ensure the standard of the functionality was adequate for users' needs. The test subjects took on the role of students within the 3D environment. They consisted of 7 postgraduate students, one secondary school teacher, and one college lecturer. The user trial was set up to ensure each user was exposed to all the features of CLEV-R. The test subjects registered for the system and choose an avatar to represent them in the 3D environment. Prior to the trial, each test subject received an image and a PowerPoint file via e-mail. They were also supplied with instructions for completing the user trial and an evaluation sheet. At an appointed time, those taking part in the trail attended a synchronous lecture in which a tutor presented slides and gave instructions on how to use CLEV-R. After a short class, the students were asked to complete a set of tasks, which involved exploring the 3D environment and the features it provides. For example, they were asked to access a set of notes from the library, view them, and download them to their own computer. Other tasks included uploading the supplied image. The participants were also required to test the text and audio communication features. Toward the end of the trial, all test subjects were instructed to attend a virtual meeting room where they had to upload their PowerPoint slides and discuss them. By assigning tasks in this way, each student experienced the facilities provided in CLEV-R and was able to supply feedback on both usability and the usefulness of the system for learning.

The results were encouraging and all test subjects were enthusiastic about the system.

As intuitive navigation within the 3D environment is a key aspect of the system, we were particularly interested in user-feedback on this matter. The feedback relating to navigation was mixed. While those who had prior experience of using 3D environments found manoeuvring easy, it took novice users some time to become familiar with the controls. Entering rooms proved particularly difficult and so we are improving the situation by removing doors and making the doorways wider. Test subjects found the communication controls easy to use, although some experienced an echo while using the audio controls. This can occur if speakers are too close in proximity to the microphone and so clearer instructions on this could resolve this issue. The lecture room was seen as an effective medium for teaching, and the participants particularly liked the real-time communication features. All users successfully uploaded files to the CLEV-R environment and collaborated on a task. Another key area of interest to us during this evaluation was the users' sense of immersion and presence within the 3D learning environment. Most of the test users felt part of a group and no one felt isolated during the evaluation. All of the subjects were engaged in the interactive environment and their interest in learning was maintained throughout the trial.

Test users with previous experience of e-learning systems were asked to comment further on CLEV-R, comparing its features to the e-learning systems previously encountered. The collaborative features of CLEV-R proved popular with these test subjects and they also liked their awareness of others during the online session. This is a feature they found lacking in other e-learning systems. Since the user trial, we have begun to address some of the technical issues that arose. We are also using the comments and feedback from the test subjects to improve the set of features CLEV-R provides. This preliminary user-trial paves the way for a more extensive trial with a larger number of test users in the near future.

mCLEV-R

We are developing a mobile version of CLEV-R that will provide "anytime-anywhere" access to learning resources. This mobile version is called mCLEV-R and provides the opportunity for people

on the move to work through course material and to communicate with course tutors and other users in real time when they cannot be stationed at a fixed location. Our research in this field has focused on the following aspects:

- Exploring the use of a 3D interface for m-learning.
- Examining the technical capabilities of small mobile devices with regard to 3D graphics. In particular, we are exploring the use of the Virtual Reality Modelling Language (VRML) on personal digital assistants (PDAs).
- Facilitating the use of PDAs as a collaboration tool for e-learning.
- Evaluating this ubiquitous and pervasive system for learning.

Mobile devices have certain limitations when it comes to developing any application for them. Small screen sizes with low resolution make it difficult to ensure an application looks well and displays all necessary information. Limited memory for running applications, and also for storage, means applications need to be light weight. Also, lack of software support is another concern. Their inexpensive and convenient nature, however, makes them the ideal target for a learning application, and once mCLEV-R has been developed, we feel it will be of great benefit to any mobile student.

Due to the device limitations mentioned above, the system needs to be greatly modified for this new platform. Firstly, the 3D environment provided in mCLEV-R is much simpler than that of the full-scale system. Only features absolutely necessary are downloaded to mobile devices, and even then may need to be simplified for more efficient rendering. For example, textures can be replaced with simple colours and complex shapes can be simplified or removed altogether depending on their importance in the environment. It is also necessary to reduce the functionality for the mobile system. Thus, we must prioritise the features of the system, carefully selecting those best suited to a mobile platform. The two most important features of our system are firstly to present learning content to users, and secondly to support social interaction among connected users. Thus mCLEV-R supports both these features.

Access to course content is supported through synchronisation with a desktop PC and by download from the 3D interface on the mobile device

Figure 10. The mCLEV-R interface on PDA

(see Figure 10). Due to small screen size and low resolution of the PDAs, course notes cannot be displayed clearly within the 3D environment. Therefore, we use external applications such as Pocket Word, Pocket Slideshow, and Pocket Adobe Acrobat Reader to open course files on these devices. Technological limitations of these devices, including lack of software support, mean it is unfortunately not possible to ensure full consistency with the desktop version of CLEV-R. Thus students on mobile devices are not aware when a tutor changes a lecture slide and are not updated about other users' locations in the environment. They can, however, be notified about important changes in the environment through the communication techniques provided. Both text chat, as seen in Figure 7, and audio chat are available in the mobile system. Thus, mCLEV-R users are continually aware of other connected users and can converse via these communication modes in real time, making them feel part of the learning community.

We have introduced our system, CLEV-R, a collaborative learning environment with virtual reality that is used for e-learning. It takes a novel and modern approach to address some of the issues and problems with existing e-learning systems. A 3D university is provided where students can learn together with learning material delivered through the medium of VR. Online lectures are enhanced through the addition of multimedia, animations, and live video. These enhancements help to stimulate and motivate students. Recognising the importance of social learning, CLEV-R provides a social setting where students can interact with one another. Collaboration on group and project work is also possible using designated areas and special tools within the environment. CLEV-R promises to improve the usability of online learning and enhance the students' learning experience. The extension of the system to mobile devices is highly innovative and presents interesting research challenges. A subset of the functionality is provided on PDAs, and we feel mCLEV-R will be an invaluable accompaniment to the full-scale system, giving students the opportunities an ubiquitous learning environment provides. Once development is complete, we look forward to the results of an extensive evaluation determining their true value for online learning.

FUTURE TRENDS FOR COLLABORATIVE LEARNING ENVIRONMENTS WITH VIRTUAL REALITY

The e-learning domain is set to increase, as the focus in society shifts to life-long learning. People of all ages are now partaking in education courses, older generations are returning to the classroom to learn new skill sets, and younger generations are staying in education longer. E-learning is an ideal solution for the needs of life-long learning, allowing people to access course content, material, and help where and when they want. There is thus no doubt that research into the provision of e-learning courses will continue well into the future. But what exactly does the future have in store for e-learning?

Throughout this chapter, we have examined the use of collaborative virtual reality environments for online learning. 3D interfaces like this are already being developed in other domains, such as e-commerce and tourism, and we see them becoming much more widespread in the future. 3D environments for online education are very effective, as they are stimulating and motivating and so engage the student in their learning. They are made all the more amenable and inviting when combined with the multi-user and collaborative features discussed above. As more of these systems emerge, the advantages (such as better retention rates for online courses and more interaction, discussion, and cooperation between students) will be seen by educators. We feel this will ultimately lead to a general accep-

tance of VR and 3D as a medium for learning. Of course, it would be naïve for us to think that this concludes research into e-learning. E-learning and distance learning are continually evolving, adapting to new technologies and new requirements. We have no doubt that this will continue in the future and so, based on current state of the art technologies, we now surmise what the next milestones in e-learning will be.

There is great potential within these 3D learning environments to further enhance a user's learning experience. Firstly, the environment could be personalised to each user's individual preferences, thus making the environment even more welcoming to them. Personalisation and adaptive user interfaces are areas of high interest at the moment (Liu, Wong, & Hui, 2003; Ye & Herbert, 2004) and they could also be applied to these 3D learning environments. These techniques examine user profiles and preferences, and subsequently adjust system features for each user accordingly. Therefore, a user could alter the physical appearance of the virtual world and state how they wish course notes to be displayed, what format they want them to be in for download, whether they want them to appear online to other users, and so forth. Software agents could also be added to the system to remove some mundane tasks for the user and oversee the management of the environment (McArdle, Monahan, Bertolotto, & Mangina, 2005). They could be used to make recommendations to the users about course material and other similar users, and indeed to help tutors keep track of students' progress and results.

While the use of VR on desktop computers continues to grow in popularity, research is now being carried out on the possible uses of augmented reality (AR). This branch of computer science is concerned with augmenting a real world environment or scene with computer-generated objects. This is often achieved through the use of head mounted displays (HMDs); the user wears a special pair of glasses and as they look at objects in the real world, computer generated 3D objects and data are superimposed into their view via the glasses. While AR research has been taking place for some time, it is only now that hardware is able to deliver results which are satisfactory to the user. One of the driving forces behind this technology is the area of medical visualisation. An example of a system using a HMD to project a 3D model of a patient's liver is described by Bornik, Beichel, Reitinger, Sorantin, Werkgartner, Leberl, et al. (2003). This tool renders the patient's liver from an x-ray computed tomography (CT) scan and enables surgeons to measure the dimensions of the liver to locate a tumour prior to surgery. The surgeon can then manipulate this model to obtain different viewing angles and see features in more detail. One drawback of using head mounted displays is that they tend to be cumbersome to wear and do not feel very natural. Again, it is in the medical visualisation arena in which strides are being made to alleviate this issue. Schnaider, Schwald, Seibert, and Weller (2003) have developed MEDARPA, a tool to assist with minimal invasive surgery. It consists of the practitioner looking through a screen, which is placed over the patient. Based on information from previous scans, an internal view of the patient is displayed on the screen; sensors track the position of the screen and the doctor's tools and update the image accordingly. The doctor can then use the on screen image as a guide for the keyhole surgery.

These two examples from within the medical domain show where this technology is heading. As mentioned earlier, distance learning and learning in general, has always evolved with technology and there is no reason why it will not embrace this new AR technology. One can easily see how this technology could be adapted for education. The very fact that 3D models themselves can be projected into the real world provide a means for students to see for themselves things that may have been dealt with in a theoretical way within

the traditional classroom. For example, complex chemical structures, human organs, computer components, and sophisticated machinery can be projected into the real world for students to interact with and manipulate, therefore increasing their understanding of a topic. We see the future of collaborative learning environments as discussed above adapting to augmented reality. A student, or indeed a teacher who is not able to attend the physical classroom setting, may have their representation projected into the classroom. They can then see the lecture taking place, and others are aware of their presence and can interact with them in a natural way. An alternative to this idea is for a holographic representation of the teacher to be projected to the location where the remote student is, in a similar style to that seen in the Star Wars movie series. Unfortunately, acceptance and widespread availability of this form of technology is a long way off.

Before the advent of AR technologies that offer truly ubiquitous and pervasive learning environments, the use of mobile computers as a supplement to e-learning will increase. As people lead busier and more hectic lives, the need to access learning content while on the move will become paramount, and m-learning will emerge as a solution to this. We envisage great improvements in mobile technologies that will allow people to access vast amounts of learning content from PDAs and mobile phones in the future. Improvements and growth in wireless networks will allow more sophisticated communication and collaborative techniques, and will also make it possible for mobile users to download large detailed virtual environments and fully partake in synchronous and interactive learning scenarios like the one CLEV-R permits. Systems like ours, with its collaborative tools supporting interaction between students, will be particularly beneficial and will play an important role in moulding the future of m-learning.

CONCLUSION

This chapter gives a brief insight into the history of distance learning, outlining how it is continually evolving and adapting to new technologies, and arguing that e-learning will embrace the current range of VR technologies now available. We particularly focus on the need for collaboration within e-learning systems. This chapter shows how 3D graphics, with its roots as a modelling tool for engineers, has been used in the past for computer games, and outlines how it is being used today in activities such as urban planning, tourism, and e-commerce. Online collaboration tools initially grew out of text-based fantasy computer games, and this chapter charts how they evolved, becoming conference tools and later acting as social aids. The use of 3D, combined with collaboration techniques, is a more recent phenomenon and several examples of such systems being developed have been discussed. These systems offer a new form of e-learning, addressing many of the issues, such as isolation and lack of motivation, which students often experience while using text-based e-learning environments. The benefits and weaknesses of these new VR e-learning systems are presented.

This chapter describes our efforts in developing a VR e-learning system called CLEV-R. Like some of the other systems, CLEV-R has the remit of providing a motivating and stimulating multi-user 3D environment; however, our research recognises the importance of social learning within groups of students, and so offers specific features to facilitate this. CLEV-R is an intuitive, multimedia rich Web-based environment, which can be used both as a complete distance learning solution and as an accompaniment to traditional classroom-based teaching. Unlike the other systems presented in this chapter, CLEV-R uniquely offers support for providing learning material and collaboration tools on mobile devices, such as

PDAs. This addition provides anytime, anywhere access to course material and allows students to interact while they are away from their desktop computer. The use of mobile devices is a new avenue in the e-learning paradigm, which has recently been termed m-learning. As the need for pervasive learning environments comes to the forefront of the research community, the use of m-learning is sure to increase. We have also discussed some interesting future trends within 3D visualisations, particularly demonstrating how the need for improvements in medical visualisation is fuelling research in Augmented Reality. We conclude this chapter by proposing how AR can be adapted for use as a tool within e-learning to provide a truly pervasive and ubiquitous learning environment.

REFERENCES

Bornik, A., Beichel, R., Reitinger, B., Sorantin, E., Werkgartner, G., Leberl, F., et al. (2003). Augmented reality based liver surgery planning. *Computer Graphics Forum, 22*(4), 795-796.

Bouras, C., Fotakis, D., Kapoulas, V., Koubek, A., Mayer, H., & Rehatscheck, H. (1999, June 7-11). In *Proceedings of the Virtual European School-VES, IEEE Multimedia Systems'99, Special Session on European Projects,* Florence, Italy (pp. 1055-1057).

Bouras, C., Philopoulos, A., & Tsiatsos, T. (2001, July). E-learning through distributed virtual environments. *Journal of Network and Computer Applications, 24*(3), 175-199.

Bouras, C., Triantafillou, V., & Tsiastsos, T. (2001, June 25-30). Aspects of collaborative environments using distributed virtual environments. In *Proceedings of the ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 173-178). Tampre, Finland.

Bouras, C., & Tsiatsos, T. (2006, June). Educational virtual environments: Design rationale and architecture. *Multimedia tools and applications, 29*(2), 153-173.

Burigat, S., & Chittaro, L. (2005, April). Location-aware visualization of VRML models in GPS-based mobile guides. In *Proceedings of the Web3D 2005: The 10th International Conference on 3D Web Technology* (pp. 57-64). New York.

Casher, O., Leach, C., Page, C., & Rzepa, H. (1998). Virtual reality modelling language (VRML) in Chemistry. *Chemistry in Britain* (pp 34-26).

Chee, Y.S. (2001). Networked virtual environments for collaborative learning. In *Proceedings of the Ninth International Conference on Computers in Education (ICCE/SchoolNet)* (pp. 3-11), Seoul, South Korea.

Chee, Y.S., & Hooi, C.M. (2002). C-VISions: Socialized learning through collaborative, virtual, interactive simulations. In *Proceedings of the Conference on Computer Support for Collaborative Learning (CSCL)* (pp. 687-696), Boulder, Colorado.

Chittaro, L., & Ranon, R. (2000). Virtual reality stores for 1-to-1 commerce. In *Proceedings of the CHI2000 Workshop on Designing Interactive Systems for 1-to-1 E-Commerce,* The Hague, The Netherlands.

Chittaro, L., & Ranon, R. (2002, May). New directions for the design of virtual reality interfaces to e-commerce sites. In *Proceedings of the AVI 2002: 5th International Conference on Advanced Visual Interfaces* (pp. 308-315). New York: ACM Press.

Csete, J., Wong, Y.H., & Vogel, D. (2004). Mobile devices in and out of the classroom. In *Proceedings of the 16th World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications,* Lugano, Switzerland (pp. 4729-4736).

Daku, B.L.F., & Jeffrey, K. (2000, October 18-21). *An interactive computer-based tutorial for MATLAB.* In *Proceedings of the 30th ASEE/IEEE Frontiers in Education Conference* (pp. F2D:2-F2D:7). Kansas City, Missouri.

Dalgarno, B. (2002). The potential of 3D virtual learning environments: A constructivist analysis. *Electronic Journal of Instructional Science and Technology, 5*(2).

Dickey, M.D. (2003). 3D Virtual worlds: An emerging technology for traditional and distance learning. In *Proceedings of the Ohio Learning Network; The Convergence of Learning and Technology – Windows on the Future.*

Göth, C., Häss, U.P., & Schwabe, G. (2004). Requirements for mobile learning games shown on a mobile game prototype. *Mobile Learning Anytime Everywhere,* 95-100. Learning and Skills development agency (LSDA).

Hamill, J., & O'Sullivan, C. (2003, February). Virtual Dublin – A framework for real-time urban simulation. *Journal of the Winter School of Computer Graphics, 11,* 221-225.

Hudson-Smith, A. (2002, January). 30 days in active worlds – Community, design and terrorism in a virtual world. In *The social life of avatars.* Schroeder, Springer-Verlag.

Isbell, C.L., Jr., Kearns, M., Kormann D., Singh, S., & Stone, P. (2001, July 30-August 3). Cobot in LambdaMOO: A social statistics agent. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence AAAI 2000* (pp. 36-41). Austin, Texas.

Johnson, C.A., Delhagen, K., & Yuen, E.H. (2003, July 25). H*ighlight: US e-commerce hits $230 billion in 2008.* (Business View Brief). Retrieved October 11, 2006, from Forester Research Incorporated at http://www.forrester.com

Ketamo, H. (2002). mLearning for kindergarten's mathematics teaching. In *Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education* (pp. 167-170). Vaxjo, Sweden.

Kray, C., Elting, C., Laakso, K., & Coors, V. (2003). Presenting route instructions on mobile devices. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (pp. 117-124). Miami, Florida.

Laister, J., & Kober, S. (2002). Social aspects of collaborative learning in virtual learning environments. In *Proceedings of the Networked Learning Conference,* Sheffield, UK.

Lepouosas, G., Charitos, D., Vassilakis, C., Charissi, A., & Halatsi, L. (2001, May 16-18). Building a VR museum in a mueseum. In *Proceedings of Virtual Reality International Conference,* Laval Virtual, France.

Lipman, R.R. (2002, September 23-25). Mobile 3D visualization for construction. In *Proceedings of the 19th International Symposium on Automation and Robotics in Construction* (pp. 53-58). Gaithersburg, Maryland.

Liu, J.,Wong, C.K., & Hui, K.K. (2003). An adaptive user interface based on personalized learning intelligent systems. *IEEE Intelligent Systems, 18*(2), 52-57.

Luchini, K., Quintana, C., & Soloway, E. (2003, April 5-10). Pocket PiCoMap: A case study in designing and assessing a handheld concept mapping tool for learners. In *Proceedings of the ACM Computer-Human Interaction 2003, Human Factors in Computing Systems Conference* (pp. 321-328). Ft. Lauderdale, Florida.

McArdle, G., Monahan, T., Bertolotto, M., & Mangina, E. (2005). Analysis and design of conceptual agent models for a virtual reality e-learning environment. *International Journal on Advanced Technology for Learning, 2*(3), 167-177.

Monahan, T., McArdle, G., & Bertolotto, M. (2005, August 29-September 2). Using 3D graphics for

learning and collaborating online. In *Proceedings of Eurographics 2005: Education Papers* (pp. 33-40). Dublin, Ireland.

Monahan, T., McArdle, G., Bertolotto, M., & Mangina, E. (2005, June 27- July 2). 3D user interfaces and multimedia in e-learning. In *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2005),* Montreal, Canada.

Nijholt, A. (2000). Agent-supported cooperative learning environments. In *Proceedings of the International Workshop on Advanced Learning Technologies* (pp. 17-18). Palmerston North, New Zealand.

O'Hare, G.M.P., & O'Grady, M.J. (2003). Gulliver's genie: A multi-agent system for ubiquitous and intelligent content delivery. *Computer Communications, 26*(11), 1177-1187.

Oliver, B., & Wright, F. (2003). E-learning to m-learning: What are the implications and possibilities for using mobile computing in and beyond the traditional classroom? In *Proceedings of the 4th International Conference on Information Communication Technologies in Education,* Samos, Greece.

Open University. (2005). *Media relations, fact sheet series, history of the open university.* Retrieved October 11, 2006, from http://www3.open.ac.uk/media/factsheets

O'Sullivan, C., Cassell, J., Vilhjálmsson, H., Dobbyn, S., Peters, C., Leeson W., et al. (2002). Crowd and group simulation with levels of detail for geometry, motion and behaviour. In *Proceedings of the Third Irish Workshop on Computer Graphics* (pp. 15-20).

Papagiannakis, G., Hoste, G.L., Foni, A., & Magnenat-Thalmann, N. (2001, October 25-27). Real-time photo realistic simulation of complex heritage edifices. In *Proceedings of the 7th International Conference on Virtual Systems and Multimedia VSMM01* (pp. 218-227). Berkeley, California.

Papagiannakis, G., Schertenleib, S., O'Kennedy, B., Poizat, M., Magnenat-Thalmann, N., Stoddart, A., et al. (2005, February). Mixing virtual and real scenes in the site of ancient Pompeii. *Computer Animation and Virtual Worlds, 16*(1), 11-24.

Raghupathiy, L., Grisoniz, L., Faurey, F., Marchalz, D., Caniy, M., & Chaillouz, C., (2004). An intestinal surgery simulator: Real-time collision processing and visualization. *IEEE Transactions on Visualization and Computer Graphics, 10*(6), 708-718.

Rakkolainen, I., & Vainio, T. (2001). A 3D city info for mobile users. *Computers & Graphics (Special Issue on Multimedia Appliances), 25*(4), 619-625.

Redfern, S., & Naughton, N. (2002). Collaborative virtual environments to support communication and community in Internet-based distance education. In *Proceedings of the Informing Science and IT Education, Joint International Conference* (pp. 1317-1327). Cork, Ireland.

Rickel, J., & Johnson, W.L. (1997). Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the First International Conference on Autonomous Agents* (pp. 30-38). California.

Rickel, J., & Johnson, W.L. (1999). Virtual humans for team training in virtual reality. In *Proceedings of the Ninth International Conference on AI in Education* (pp. 578-585).

Riedl, R., Barrett, T., Rowe, J., Vinson, W., & Walker, S. (2001). Sequence independent structure in distance learning. In *Proceedings of Society for Information Technology and Teacher Education INternational Conference* (pp. 1191-1193)

Rischbeck, T., & Watson, P. (2002, March 24-28). A scalable, multi-user VRML server. In *Proceedings of the IEEE Virtual Reality Conference* (pp. 199-207). Orlando, Florida.

Ryan, J., O'Sullivan, C., Bell, C., & Mooney, R. (2004). A virtual reality electrocardiography teaching tool. In *Proceeding of the Second International Conference in Biomedical Engineering* (pp. 250-253), Innsbruck, Austria.

Schilling, A., & Coors, V. (2003, Septmeber). 3D maps on mobile devices. In *Proceedings from the Design Kartenbasierter Mobiler Dienste Workshop,* Stuttgart, Germany.

Schnaider, M., Schwald, B., Seibert, H., & Weller, T. (2003). MEDARPA - An augmented reality system for supporting minimally invasive interventions. In *Proceedings of Medicine Meets Virtual Reality 2003* (pp. 312-314). Amsterdam, The Netherlands.

UNESCO (2002). *Open and distance learning, trends policy and strategy consideration. United Nations Educational Scientific and Cultural Organisation (UNESCO)* Report 2002. Retrieved October 11, 2006, from http://unesdoc.unesco.org/images/0012/001284/128463e.pdf

Wikipedia Blackboard Incorporated. *In The Wikipedia Encyclopedia*. Retrieved October 11, 2006, from http://en.wikipedia.org/wiki/Blackboard_Inc

Ye, J.H., & Herbert, J. (2004, June 28-29). Framework for user interface adaptation. In *Proceedings from the 8th ERCIM Workshop on User Interfaces for All* (vol. 3196, pp. 167-174). Vienna, Austria: Springer Verlag.

Zara, J., Benes, B., & Rodarte, R.R. (2004, September 20-24). Virtual campeche: A Web based virtual three-dimensional tour. In *Proceeding of the 5th Mexican International Conference in Computer Science*, (pp. 133-140). Colima, Mexico.

Zimmerman, G., Barnes, J., & Leventhal, L.M. (2003). A comparison of the usability and effectiveness of Web-based delivery of instructions for inherently-3D construction tasks on handheld and desktop computers. In *Proceedings of Web3D 2003* (pp. 49-54). Saint Malo, France.

APPENDIX I: INTERNET SESSION

C-VISions: Collaborative Virtual Interactive Simulations

The C-VISions research group develop interactive simulations to help students learn (*http://yamsanchee.myplace.nie.edu.sg/NUSprojects/cvisions/cvisions.htm*). The Web site above provides details of their work along with relevant publications. Use this information to prepare a presentation, outlining the background to the research along with a synopsis of the systems they have developed.

APPENDIX II: CASE STUDY

A university has been offering a virtual reality learning environment as an accompaniment to classes and as a distance learning solution for three years now. Students' acceptance of the technology has been high; however, faculty have been slow to adopt this new method of teaching.

Comp 4015 is a software engineering module offered by the university. The course involves the tutor giving a number of lectures detailing best practice methods for Java Programming; this generally takes the form of a PowerPoint presentation showing examples of poor coding. The tutor then asks individual pupils how they would fix the problem. This creates interaction and discussion within the class. Another aspect of the Comp 4015 module involves students working together on a group project, where each team must design a program to address a fictional company's needs. At the end of the course they must present their work to the class. The tutor is very reluctant to offer this course via the virtual reality environment. He feels the dialog in which the students engage in during the actual lectures will be lost. He is particularly worried that the group project will no longer be possible and it may have to become a project for individual students instead. Thirty percent of a student's final grade for this module comes from the final presentation, which students give, and the tutor is concerned that the student's presentations will no longer be possible if the module is offered via the virtual environment.

Taking the CLEV-R system described above, encourage the tutor to offer the Comp 4015 module in the virtual environment by offering advice on the following points and questions raised by the tutor:

1. The tutor has been teaching this module for many years and has all the lecture slides and material ready. He does not want to change the material in order to tailor it for the virtual environment.
2. How can the interaction, which his classes are well known for, be maintained when the module is offered in the virtual reality environment?
3. Can people who may never meet really partake in a group project and give a presentation at the end? What tools support this?
4. Students will just be anonymous, with no personality and no way for them to be distinguished or to get to know each other. Is there any way to address this?
5. Suggest how mCLEV-R, the mobile accompaniment to CLEV-R could be introduced and used on this course.

APPENDIX III: USEFUL LINKS

Human Computer Interaction Laboratory
http://hcilab.uniud.it/

MIRALab
http://www.miralab.unige.ch/

Research Unit 6
http://ru6.cti.gr/ru6/

M-Learning World
http://www.mlearningworld.com/

MOBIIearn
http://www.mobilearn.org/

TECFA
http://tecfa.unige.ch/

Augmented Reality
http://www.uni-weimar.de/~bimber/research.php

APPENDIX IV: FURTHER READING

Adelstein F., Gupta S., Richard G., III, & Schwiebert, L. (2004). *Fundamentals of mobile and pervasive computing.* McGraw-Hill Professional.

Bimber, O., & Raskar, R. (2005). *Spatial augmented reality: Merging real and virtual worlds.* A.K. Peters, Ltd.

Bowman, D.A., Kruijff, E., LaViola, J.J., & Poupyrev, I. (2004). *3D User interfaces: Theory and practice*. Addison-Wesley Professional.

Burdea, G.C., & Coiffer, P. (2003). *Virtual reality technology* (2nd ed.). Wiley-IEEE Press.

Comeaux, P. (2002). *Communication and collaboration in the online classroom: Examples and applications.* Anker Pub Co.

Mahgoub, I., & Ilyas, M. (2004). *Mobile computing handbook*. CRC Press.

McLennan, H. (1999). *Virtual reality: Case studies in design for collaboration and learning.* Information Today Inc.

Palloff, R.M., & Pratt, K. (2004). *Collaborating online: Learning together in community.* Jossey-Bass guides to online teaching and learning. Jossey-Bass.

Sherman, W.R., & Craig, A. (2002). *Understanding virtual reality: Interface, application and design.* The Morgan Kaufmann series in computer graphics. Morgan Kaufmann.

APPENDIX V: POSSIBLE PAPER TITLES/ ESSAYS

- Issues with traditional text-based e-learning systems
- Combining collaborative tools and virtual reality
- Embracing new technologies to deliver learning material over the Internet
- Mobile technologies to offer ubiquitous learning environments
- Augmented Reality: The future for education?

This work was previously published in Ubiquitous and Pervasive Knowledge and Learning Management: Semantics, Social Networking and New Media to Their Full Potential, edited by M. Lytras and A. Naeve, pp. 118-157, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.11
Empirical Validation of a Multimedia Construct for Learning

Paul Kawachi
Kurume Shin-Ai Women's College, Japan

ABSTRACT

A multimedia construct for learning based on the Theory of Transactional Distance has been developed consisting of four stages of decreasing transactional distance. This model has been applied in various teaching and learning contexts, on- and off-line, and its validation was investigated. Results confirmed in practice the four distinct sequential stages. Difficulties were discovered in navigating through the collaborative second and third stages, consistent with findings from related studies on acquiring critical thinking skills. Specific areas for attention were identified to promote learning using multimedia.

INTRODUCTION

Previous Models of Learning

Two significant models have been proposed to identify the essential steps of learning critical-thinking skills: one by Dewey (1933) and another by Brookfield (1987). Dewey proposed five phases of reflective or critical thinking:

1. Suggestions, in which the mind leaps forward to a possible solution
2. An intellectualization of the difficulty or perplexity that has been felt (directly experienced) into a problem to be solved, a question for which the answer must be sought
3. The use of one suggestion after another as a leading idea, or hypothesis, to initiate and guide observation and other operations in collection of factual material
4. The mental elaboration of the idea or supposition (reasoning, in the sense in which reasoning is a part, not the whole, of inference)
5. Testing the hypothesis by overt or imaginative action

Brookfield also proposed five phases to develop critical thinking:

1. A triggering event
2. An appraisal of the situation
3. An exploration to explain anomalies or discrepancies
4. Developing alternative perspectives
5. Integration of alternatives in ways of thinking or living

However, the steps given in the above models do not correlate with each other. The steps are not clearly distinguishable, and the actual process need not be sequenced linearly. So these models are not sufficiently clear to constitute the basis of a syllabus. A new clear and practical model is proposed based on the distinct ways of learning. And this new model will constitute the basis for an intelligent syllabus for acquiring critical-thinking skills using multimedia.

The Distinct Ways of Learning

There are four distinct ways of learning (Kawachi, 2003a): learning alone independently, alone individually, in a group cooperatively, and in a group collaboratively. Here it is important to distinguish cooperative learning from collaborative learning, in order to deploy these in the new model detailed below.

Cooperative learning essentially involves at least one member of the group who "knows" the content soon to be learned by the other(s). Learning takes place through the "knower" repeating, reiterating, recapitulating, paraphrasing, summarizing, reorganizing, or translating the point to be learned.

Collaborative learning follows a scientific process of testing out hypotheses. A participant publicly articulates his (or her) own opinion as a hypothesis, and being open to the value of conflict allows this to be negated if possible by others, in which case the original participant or another offers up a modified or alternative hypothesis for public scrutiny. In collaborative learning, disagreement and intellectual conflict are desirable interactions. All participants share in coconstructing the new knowledge together, and this learning occurs inside the group as a type of consensus achieved through analysis and argument. In collaborative learning, there was no "knower" prior to the learning process taking place (in contrast to the situation of cooperative learning).

Need for a New Model of Learning

Largely as a result of the rapid expansion of open and distance education, learning theory has undergone a revolution to a social constructivist paradigm based on cognitive concepts of how we learn. Previous models of learning have been too vague for applying to current learning practices through computer-mediated communications. Hence, there is a need for a new practical model.

NEW MULTIMEDIA LEARNING MODEL

A new model for learning critical thinking using multimedia has been proposed by Kawachi (2003b). Design is a key characteristic generally lacking in the current applications to date of computer-mediated communications adopted in conventional face-to-face or distance education courses. The presented Design for Multimedia in Learning (DML) model translates conventional theoretical models of learning into an efficient practical design for use in the multimedia educational environment. While the two leading previous models have variously postulated five phases to critical thinking for learning, this new model has four distinct stages, and is directly underpinned by Moore's (1993) Theory of Transactional Distance. This theory, which involves educative-dialogue (D), prescribed structure (S), and student autonomy (A), tries to measure the psychological distance between the student and the information

to be learned, and has been widely accepted as an effective theory underlying and informing open and distance education. The original theory only deals with one student, learning content with the interactions of a tutor. So it is adapted here to bring into account the important interactions among the student and other students (for a discussion here, see Kawachi, 2003b).

The four stages of the new DML model are as follows:

- In **Stage 1**, learning occurs in a group cooperatively, gathering and sharing information and fostering a learning community. Here synchronous-mode computer-mediated communications are best, such as chat and conferencing. However, it should not be forgotten that bridging telephony can simultaneously link 50 students synchronously with the tutor(s). Videorecording the interactions here could provide material for reflection in Stage 2, or as is often the case, the tutor as observer could take written notes for later distribution as a summary or transcription to the participants. This stage can be characterized by self-introductions (as a prelude to being a source of content material to other students), brainstorming (limited at Stage 1 to only accumulating new ideas, yet to be argued in Stage 2), involving divergent thinking to gather various different perceptions in order to explore and to frame each student's context, and helping each other as equals with obtaining content, especially in sharing personal experiences and past literature that has been read, which constitute old foundational knowledge. (Brainstorming is initiated by providing an ill-defined scenario or case study to elicit multiple perspectives.) The transactional distance initially is at a maximum (D- S-) with no teaching-dialogue and with no pre-set structure.
- In **Stage 2**, lateral-thinking (creative thinking around the problem) is used to generate and develop metaphors (an idea or conception that is basically dissimilar but formed from noting similarities between the initial information and the new concept) or new ideas, and these supported by argument. Students discuss, for example, their own problems that they have found which have brought them to participate in the current course, and then argue to identify possible solutions to each other's problems. Creative thinking here may derive from combining seemingly disparate parts, especially ideas contributed from others in different contexts into a new synergic whole. The teacher is still keeping academically at a distance away from the content under discussion, while the students are making their efforts to achieve some pre-set goals (to present own problem and reasons for engaging the current course, for example), which gives structure to their discussions (D- S+). Some time is needed for reflection here, and asynchronous modes such as e-mail and a bulletin board are effective because of the time interval incurrent between receiving the stimulus and the student's response. Moreover, these modes of interaction through written text also provide a written record to the student that enables recapitulation, retrieval of a theme, and recovery of someone's perspective, and so foster reflection.
- In **Stage 3**, the tutor engages the students with guiding comments in what Holmberg (1983) has described as a Guided Didactic Conversation, helping the students achieve the course structural requirements of understanding the general concepts to be learned (D+ S+). The tutor poses questions, and students defend their formulations. This stage is characterized by hypotheses testing and logical straightforward thinking (termed "vertical" thinking in contrast to "lateral" thinking) associated with problem solving and is collaborative. Problem-based learn-

ing can involve holding multiple alternative hypotheses at the same time, and evidence gathered can be assigned to examine simultaneously the various hypotheses. Asynchronous mode is ideal here, to allow sufficient time for cognitive connections and co-construction of new nonfoundational knowledge.

- In **Stage 4**, the final stage, the course requirements have largely been already achieved and there is no structure left, except to disseminate the achieved mental ideas and test them out in real life. This stage is characterized by experiential learning and is cooperative, and at minimum transactional distance (D+ S-), in synchronous mode, and with teaching dialogue to assist the students to reflect on their studies.

STUDENT AUTONOMY IN LEARNING

Definitions of "autonomy" in learning have in common an emphasis on the capacity to think rationally, reflect, analyze evidence, and make judgments; to know oneself and be free to form and express one's own opinions; and finally, to be able to act in the world (Tennant & Pogson, 1995). These qualities characterize the collaborative thought processes of Stage 3, and also the experiential aspect of Stage 4. Stage 1 has maximal transactional distance, and for a student to succeed here in independent learning, Moore (1993) pointed out that the student would need maximum autonomy (p. 27). Autonomy is thus seen as a highly powerful and desirable quality for independent learners. Not all students bring this high level of autonomy with them initially into their studies, and so the tutor must bring the student around to acquire this autonomy. The DML model illustrates a cyclical process—even an iterative process—through Stages 1 to 4 to equip and bring the student to go onto independent learning in a further new cycle starting at Stage 1 in a new learning venture.

Autonomy has also been related to recognizing one's interdependence on others (Boud, 1988). Interdependence relates to understanding the need to learn together with others, either in cooperative mode or at other times in collaborative. Interdependence is a maturity characterizing an adult student and is acquired through awareness and prior experience of the critical-thinking process. Toward the end of Stage 4, the student can have acquired this sense of interdependence. So in entering a new Stage 1 interaction, the student may be interdependent (post-Stage 4) and once more newly independent (starting a fresh Stage 1). These attributes of independence and interdependence have already been found to be separate, orthogonal, and coexisting in mature students at the end of their course (Chen & Willits, 1999).

While autonomy is defined as an attribute of the student, different distance education programs and the different stages in the DML model relate to different levels of autonomy for the student to be a successful learner. In a program at Stage 2, the deployed structure means that the student is charged with thinking rationally, but horizontally rather than vertically, and is analyzing already-given evidence, rather than finding new evidence, so the quality of autonomy is somewhat measured to fit the limited freedom given to the student. At Stage 3, different qualities of autonomy for hypotheses testing are needed for success—including a mature openness to new ideas that might be in conflict with one's previous and present conceived view of the world. The student needs to exercise the freedom to formulate or reformulate one's own conceptions. While in Stage 4, the quality of autonomy should include the willingness and ability to act to test out these newly constructed ideas to see experientially how they operate in practice.

It is difficult, therefore, and moreover unhelpful to assign an integrated level of autonomy to each stage in the DML model. The student should utilize

measured amounts of the various qualities that constitute autonomy during each stage to support learning. Can the tutor and institution influence the level and qualities of autonomy used by the student? Yes. And explicit clear advice from the tutor may be all that is required. The student, however, might not yet possess the skills for exercising the full range of qualities constituting autonomy (in other words, is unequipped for full autonomy). The novice and nonexpert will likely need scaffolding help at different stages to cope.

SCAFFOLDING FOR LEARNING

Scaffolding is the intervention of a tutor in a process that enables the student to solve a problem, carry out a task, or achieve a goal that would be beyond the student's unassisted efforts (Wood et al., 1976, p. 190). In providing individualized scaffolding, the tutor knows the intended knowledge to be learned and has a fair grasp of the prospective development for the student. The distance between the unassisted level of capability and the potential level that can be achieved through scaffolding is Vygotsky's (1978) "zone of proximal development" (p.86). Vygotsky included the opportunity that such scaffolding could be from "more capable" other students, indicating a cooperative assistance (as opposed to a collaborative process). Wood et al. (1976) made this very clear: tutoring is "the means whereby an adult or 'expert' helps somebody who is less adult or less expert. …a situation in which one member 'knows the answer' and the other does not" (p. 89). Accordingly, we might be advised to reserve the term "tutor" for the cooperative Stages 1 and 4 only, and use a term "facilitator" for the collaborative processes of Stages 2 and 3.

In Stage 1, tutor intervention providing scaffolding includes making the outcomes of studying explicit to the student and ensuring that the student can comprehend the aims and objectives. If not, then tutor feedback and error correction become merely vehicles of information for imitation and copying, and vaporize these opportunities to acquire mastery. The tutor need not exercise full control over the discovery process. It is recognized that students also acquire learning through unexpected accidental discovery of knowledge.

Both Stage 2 (D- S+) and Stage 3 (D+ S+) are characterized by added structure.

In Stage 3, scaffolding should add a safe structure for the interactions involved in the analytic argumentation of hypotheses testing, which have led to some students feeling wounded, by so-called flaming. Zimmer (1995) proposed an effective framework involving three functional turn-taking steps ABA between two persons A and B, which when repeated as BAB give both participants the opportunities to give opinions and receive counteropinions empathetically, as follows:

A) (Hello) Affirm + Elicitation
B) Opinion + Request understanding
A) Confirm + Counteropinion
B) Affirm + Elicitation
A) Opinion + Request understanding
B) Confirm + Counteropinion

I should also like to propose another framework drawn from some ideas of Probst (1987) for collaborative learning in literature and art, in which transactions are not aimed at hypotheses-testing characterized by counteropinion but rather a new insight built on critical reflection that while shared may be personalized in each individual. In literature, learning is not cooperative: there is no "knower"; the tutor does not guide the student to some pre-set conclusion of the meaning of the text. In literature, the tutor or any student (A) elicits opinion to initiate the three functional turn-taking steps BAB (followed by ABA), as follows:

A) (Hello) Affirm + Elicitation
B) Opinion/Analysis + Request understanding
A) Affirm + Elicitation of evidence

B) Reflect + Elicit other opinions/Analyses

A) Opinion/Analysis + Request understanding
B) Affirm + Elicitation of evidence
A) Reflect + Elicit other opinions/Analyses

This framework—basically of reflective analysis followed by articulation, bring in ideas from their own reading or those elicited from other students, then repeat reflective analysis with accommodation to construct a new insight—involves the same cognitive processes that occur in individual learning. In the group, content comes from texts and other students, while in individual learning, content comes only from texts. In both cases, it is the transactions between the student and the content that creates the new knowledge in the student.

Courses based on experiential learning that focus on Stage 4 in synchronous mode can also benefit from explicit scaffolding. In non-face-to-face (nonvideo, nonaudio) synchronous "chat" text-based conferencing, students should be directed to articulate their feelings explicitly. Neubauer (2003a) found that once students had become skilled in explicitly stating their feelings (such as "I am confused…"), then their learning improved by better sharing experiences, and they then more highly valued their text-based content—more than if they had used visual face-to-face cues. So, scaffolding can also assist in Stage 4 synchronous chat experiential learning.

ON THE NUMBER OF PARTICIPANTS

In both the above frameworks, I suggest that any participant(s) may be behind either voice, so the framework could be effective for more than two persons at the same time. Bork (2001) has suggested that the optimal number may be four in collaborative transactions, in an optimal online class size of 20 students, while six has been reported by Laurillard (2002), and about 10 by others. Wang (2002) has asserted that engaging as many participants as possible would maximize diversity and optimize collaborative learning. Zimmer (1995) has found that provided participants are aware of the framework, then collaborative learning succeeded in practice for a group of 12 students.

The optimum number of active participants in synchronous cooperative learning is different from that for asynchronous collaborative learning. An online survey of those on the DEOS-L listserv, who have had relevant experience in conducting synchronous chat (Neubauer, 2003b), found the optimum number was from 10 to 20 students: if students were new to the synchronous media, then five to seven students was optimum; in groups of 10 to 15 mixed-experience students, then 10 was optimum; while if students were experienced and the moderator (tutor) also was experienced, then 15 was optimum. And the upper limit of 20 students was suggested to keep the discussion at a sufficiently fast rate to maintain high interest levels. There seemed to be a marked difference between respondents who found that five to seven was optimum and those who found that 20 was optimum, and this difference might be related to the task at hand. Five to seven new students would imply that they were at Stage 1, forming a learning community with personal introductions and so on, while 20 students were likely at Stage 4, sharing course experiences. A note should be added here to the effect that non-native-speakers of English might be slower and more apprehensive (than native speakers) about their actively participating in synchronous discussions (see, for example, Briguglio, 2000; Kawachi, 2000). That these synchronous discussions are cooperative and not collaborative, however, should mean that their state anxiety should be lower and performance higher than if collaborative discussion were conducted synchronously.

USING A FRAMEWORK

These two frameworks illustrate and scaffold the interactions, either synchronous or asynchronous, for learning collaboratively in a group. The framework indicates what content should optimally be included in an utterance, and specifies in what serial order to progress towards achieving discovery and coconstruction of new understanding and new knowledge. It should also be noted that the use of a framework also implies some timeliness in replies. The system would not function if turn-taking were violated or not forthcoming. Participants need to take responsibility for the group succeeding by actively providing what is required and when it is required. In this way, some pacing is inevitable if the group is to move towards achieving its goal.

To some large extent, nonresponse in an asynchronous environment can be overcome by others offering up the required content in time. This is often the case in synchronous free discussions. However, group cohesiveness depends on the active participation of all members of the group. If a student does not participate, the group is fragmented and not functioning as a whole. Prior to the task, coping strategies should be acquired, agreed upon, and then used when required, such as prearranging the time frame allowed within which a student should contribute, pairing up students to provide backup in case one is at a loss, or having the moderator provide behind-the-scenes coaxing and elicitation.

METHODOLOGY FOR VALIDATION OF THE MODEL

Research into preferred learning styles has suggested that while some students may be field-dependent, others are field-independent. Lyons et al. (1999) described how some are so-called right-brain dominant. These students tend to be intuitive and prefer informal unstructured learning environments and group discussions in empathetic elicitation, sharing, and valuing each other's experiences and views (who would prefer cooperative learning in a group). Others are so-called left-brain dominant and are analytic, rational, and objective (who would prefer collaborative learning in a group).

In order to validate the model empirically, hypertext linkages were added purposively into a Web-based course. The Internet is a nonnarrative media in which no predetermined pathway through it is provided to the student newly logging on. Hypertext linkages on corporate business Web sites have been categorized by Harrison (2002), but there has been no categorization to date of hypertext usage in educational Web sites. Here, some links were colored red to indicate that examples could be reached by clicking on the highlighted linkage, while other links were colored blue to indicate to the student specifically that reasons could be reached. The courseware was reduced in content by removing all preexisting or customary references to examples and reasons except for the colored hypertext links.

It was then postulated that during traveling through the courseware, some students preferred to see examples, while others preferred to see reasons, with both groups achieving learning of the general concepts with no significant difference in achieved quality of learning.

The students examined in this study were all Japanese, and Japanese students are known to prefer cooperative learning in a group and avoid critical evaluation of others, preferring instead to preserve group harmony through empathetic sharing (Kawachi, 2000).

By coloring each hypertext or telling the students directly what color it would be or which content could be reached through which link, it was then planned that students would not open a link simply from curiosity but would pass across any link they decided was not wanted and move onto opening a link that might be helpful in their accomplishing the task at hand. Students were

monitored in their selection. Students were also required to keep journals as a written "think-aloud" record for formative and summative evaluation. Students were also interviewed during and after their online studies. E-mail messages were also kept. Students were continually encouraged to interact with each other. This was to keep the group on task cohesively, providing peer support and pacing to some degree, as well as for the designed cooperative or collaborative interactions.

RESULTS

The above method for empirical validation of the DML model in Japan using specially designed hypertext courseware to investigate cooperative and collaborative pathways during learning was not entirely successful. The study found that students at the undergraduate level could successfully move through the first two stages but could not engage the third stage due to lack in sufficient foundational knowledge and experiential maturity. Interviews were conducted on the students, but these also failed to identify any cause for the breakdown in the learning cycle. Validation at the graduate and continuing adult education level is ongoing.

Course and Student Assessment

Summative records of achieved learning from each student indicated the particular stage reached by the student, and to a fair degree of accuracy, where within a stage was reached by the student. In each stage, indeed at any time throughout the course, interactions were recorded for formative and summative evaluations, of both the course itself and of the student's individual participation, contribution, learning process (including choices made), and quality of learning outcomes. In Stage 1, handwritten notes, audiorecording, or audio-video recording can serve these purposes. Only written reports, interviews, and tutor observations were used in the present study. In Stage 2 and Stage 3, the asynchronous modes are performed through written contributions, such as by mail, teletext, fax, or e-mail, so that records can be easily stored and retrieved. Nonacademic and academic exchanges between the student and others can usually be recorded (though recording telephone conversations needs informed consent). At the present time (June 2003), it remains technologically difficult, if not impossible, to record the hypertext-enabled learning narratives of each student. Some adaptive hypermedia can restrict the available hypertext choices, but at present, think-aloud, recall, and separate audio-video recording are the only means with which to track the pathways and learning processes of the students who were using interactive hypermedia. In the present study, the student journals, triangulated with interviews and observations, were used. Stage 4 is characterized by social constructivist experiential learning, which usually entails some form of public articulation of the student's tentative or summative perspective achieved from the previous stages. For example, a written thesis is the most common instrument of evaluation here. Oral presentation at a conference and publication of a report in an academic journal are also common instruments. In the present study, the final demonstration was different depending on the course. In all courses, there was a written summative report from each student. In one course, there was project work including a poster presentation and group journal thesis published. This thesis included individual reports of pathways and a group collective report.

In the empirical validation of the DML model in this study, the students were not paced, but the course was of predetermined duration. The aim at the outset was to bring all the students through all four stages to present some new personal meaning they had each achieved through the four-stage process. Observations and written records gathered during the course were revealing that many students were slower than expected—this

was even after the course was tailored to be at a comprehensible level fitting to each particular class. Within-class individual cognitive and affective differences were greater than expected. It thus transpired that the summative reports from the students, rather than confirming all had successfully completed the four stages, instead revealed the location within the model that they had each reached.

The small seminar class of six second-year undergraduates completed the four stages during the one year and adequately demonstrated their new socially constructed knowledge in an exhibition presentation, in a published journal, and in reflective reports of their experience and how the course had changed their thinking. The teaching aim was to scaffold and promote a desire in each of them for lifelong learning. Two of the six went on to engage in higher learning at another university.

Adult Motivations to Learn

In this validation study, student motivation to learn in a preferred way had a potential influence on the performance in certain stages. Therefore, to investigate any influence, students were surveyed by questionnaire on their preferred approaches to learning and their motivations. The questionnaires and self-reports were followed up by interviews. How to initiate each and all the various intrinsic motivations to learn has been previously reported by Kawachi (2002c). However, that study was based on the taxonomy of Gibbs et al. (1984), which used data from about 1960 which pre-date multimedia learning technologies. Briefly, there are the four intrinsic motivations: vocational, academic, personal, and social. These were discovered in the present study simultaneously in varying levels depending on the task and with individual differences. However, beyond these, there was suggested some motivation to lifelong learning that was difficult to accommodate within the nearest category of intrinsic personal challenge. This motivation was suggested by only the older postgraduate students. It is tentatively labeled the "aesthetic" motivation to learn. The discovery and illuminatory methodology used here was informed by various interview open responses leading into focused discussions, and it followed a grounded-theory approach. Two orthogonal dimensions were found and labeled as positive and negative incidences of *jouissance* occurring accidentally during the learning process. These incidences only occurred when the student was markedly actively learning—struggling to construct meaning to discover suddenly how things fit together in a shot of joy (positive *jouissance*) or how things had been mistaken and misunderstood (negative *jouissance*). This aesthetic motivation was concluded to be acting along the process of the interaction between the student and the content-to-be-learned (actually *to* the student *from* the content-to-be-learned, a unidirectional motivation). Aesthetic motivation derives *from* the process. There was a similar motivation acting in the opposite direction —*to* the process—of expressive motivation, in which the student is driven to proceed, by the joy of doing (as might occur for example in writing poetry, or fine-art painting). As an illustration, aesthetic motivation drives a hobby fisherman; positive *jouissance* occurs when the fisherman catches a surprisingly large fish, and alternatively, negative *jouissance* occurs when a fish escapes suddenly. Both these types occur only in the adult or mature person with an already fully formed self or culture, and they occur as a bursting of this bubble, momentarily and transiently. The fisherman's experience is increased by the *jouissance*, and he is more driven to continue fishing. Aesthetic motivation is the motivation to lifelong learning. The tutor needs to understand the limits of the student's context or worldview and guide the student to approach the limits of his or her world, hopefully to experience *jouissance* and initiate aesthetic motivation to learn.

Summary of Results

Repeated empirical studies found that only small classes with close tutor moderating, and preferably of students with sufficient background knowledge, could successfully engage the collaborative learning tasks and complete all four stages in this model. Most undergraduate students, especially in larger classes or even in small groups but with reduced tutor monitoring, could not engage the collaborative Stage 3. A similar finding was also reported by Perry (1970), in the United States, who concluded that college students were maybe not yet sufficiently mature to acquire the skills of critical thinking. It is well known (Kawachi, 2003c) that collaborative learning characterizes the construction of nonfoundational (graduate-level) knowledge rather than the acquisition of foundational knowledge (at the undergraduate level).

DISCUSSION

Scaffold Efficacy

The DML model was designed and intended to act as scaffolding to guide the teacher (in the present study, the author), inform the process, and assist student learning. It was not completely successful. This was due to the limited duration of the course and the levels of maturity in the students. The limited duration of the shorter (6-month) courses meant that the self-pacing or unpaced nature would not allow for the students to complete the full learning cycle. The low levels of maturity in the undergraduate younger students meant that they found much difficulty in navigating Stage 3.

Four separate modes of learning were serially linked in this DML model. Stage 1 employed synchronous media for cooperative brainstorming; Stage 2 employed asynchronous media for collaborative lateral thinking; Stage 3 employed asynchronous media for collaborative vertical thinking and problem-based learning; and Stage 4 employed synchronous media for experiential learning. This model indicates the need to change the type of media employed during the learning process, for example, from synchronous to asynchronous to move from Stage 1 to Stage 2. While Stage 3 proved difficult for some students, the largest hurdle was found in moving from asynchronous collaborative Stage 2 to asynchronous collaborative Stage 3. This needs to be discussed. The task activities of Stage 3 require the students to raise doubts about others, to question the teacher and the text, and to search for one's own opinion, even though this might be against the established opinions of others in authority. One reason for the students not moving into Stage 3 was that the activities of Stage 3 were inconsistent and incongruous with their own lives or cultural views of the world (for example, see Briguglio, 2000, p. 3, for a discussion of Jones, 1999, unpublished report).

These questioning skills may be characteristic of a mature adult. In support of this DML model to master these questioning skills of Stage 3, Halpern (1984) reported that all adults should learn to question input prior to acquisition, in what he described as a "content" effect: "When we reason we do not automatically accept the given premises as true. We use our knowledge about the topic (content) to judge the veracity of the premises and to supply additional information that influences which conclusion we will accept as valid" (p. 359). Adults generally have more experience than adolescents from which to draw additional information, so they can be expected to be more questioning during learning from a teacher or other resource. Younger or immature adults can be expected to not yet hold adequate foundational knowledge with which to engage the Stage 3 questioning and answering.

Moreover, a gender difference might be operating here. Raising doubts about others, and having others raise doubts about you, may be an undesirable activity for some students—not

only for women but also for those who may be disadvantaged by physical or mental dysfunctions and those who may not be adequately literate. There are many kinds of literacy involved here—linguistic literacy in a first language, linguistic literacy in the language as medium of the education (notably in English as a foreign language), information literacy (the capability to find information efficiently), cyber-literacy (the capability to handle virtual systems and manage oneself within these), and technological literacy (the capabilities to manage and interact through the human–computer interface). In such cases, the need for conservation of self may rise higher than the internal drive or need for progression through further education.

Women who try distance education may be more likely to bring with them higher levels of self-doubt and anxieties that can add a more cautious approach to their questioning of authorities. Also, traditionally, higher education has not been part of their world and self-concept, so they are operating in a somewhat alien world, and one that is not congruous with their present conception of the world. This may also be true for men, because adults generally have already established their social world and self, and where this does not include higher education—as in those who engage in higher education for the first time as a "second chance"—then the students might understandably be reluctant to argue with others in academia. Adults who are returning to higher education or are at the postgraduate level may find no incongruity.

Belenky et al. (1997) reported that the aim for participating in education is different between women and men. They write that women want to be at the center and not be far out from others, and they value the comfort that being in the group brings to their learning and self-development. They write that men, on the other hand, want to excel and be out ahead of the group, and they may feel threatened by another person being too close or approaching. Asynchronous media may provide women the time needed to move the group forward as a whole, without one moving alone, too far from the center. However, this might require unusually good communication skills and literacies.

On Pacing

There was no pacing imposed during this empirical validation of the model. Four courses were each of one 6-month academic semester, consisting of about 15 lessons, each 90 minutes, plus out-of-class interactions equivalent to at least a further 200 hours and, in many cases, much more. In another two courses, the duration was double that and continued for 1 year. No pacing was used, because through previous experience, it was found that pacing induced students to adopt a performance orientation (for discussion, see Abrami & Bures, 1996, p. 38), rather than adopt a deep approach to their learning. This was an ethical decision, which negatively confounded the findings. However, literature studies later indicated that in paced (Gunawardena et al., 1997, 2001) and unpaced (McKinnon, 1976; Piaget, 1977; Renner, 1976a, 1976b) learning, students similarly reached to various levels, not completing the four stages, and mostly reaching to somewhere between the middle of Stage 2 and the middle of Stage 3, as in the present study.

In the two courses of one-year duration, findings showed that the students generally had reached the end of Stage 4. One course was a small seminar class of six second-year undergraduate students, and the other was of eight postgraduate adult students. Both were closely guided by the tutor. A similar class of adult students was unpaced and closely guided but over only six months, and they reached only to the middle of Stage 3.

All the courses were compulsory, so no student was allowed to drop out without having to repeat the course. This fact influenced the decision to have no pacing and to focus on the students' deep learning achievements.

Other Studies Measuring Transactional Distance

The present study, through close monitoring, followed each student through the learning-cycle process from an initial maximum transactional distance to less and less transactional distance. Hypertext navigation paths and serial written reports were used, together with interviews. These were very effective as measures of the progress of each student and were fairly effective as a measure of the transactional distance. The transactional distance varied during the course, becoming more reduced. Thus, the present study was a longitudinal study of measuring transactional distance.

A cross-sectional study measuring transactional distance was recently reported by Chen and Willits (1999).

Chen and Willits (1999) designed, piloted, and applied a questionnaire to measure the transactional distance in a videoconferencing course. They applied factor analysis to determine the loadings on dialogue, structure, and autonomy. They surveyed 202 students participating in 12 different courses (suggesting that their questionnaire had up to 70 items, if the 202 were treated as one cohort together for factor analysis). The items in each factor indicated that each concept of D, S, and A was complex and not simple. Their study was limited by the fact that they could not use factor analysis to discover structure, dialogue, and autonomy as three factors initially and used simply three separate questionnaires pasted together as one, and three separate analyses – one for each of them. This was likely because dialogue, structure, and autonomy are interrelated by a simplex structure, not a hierarchical structure. Factor analysis is inappropriate for simplex structures (Bynner & Romney, 1986). Dialogue and structure are related horizontally in that the amount of structure influences the amount of dialogue, so a simplex structure exists, and factor analysis should not be used. Structural path analysis would reveal this, but Chen and Willits did not report doing any path analysis. Nevertheless, they found that lower transactional distance was correlated with a higher level of learning outcome.

Their results support the use of the DML model to reduce systematically the transactional distance during a course to increase the quality and level of learning.

IMPLICATIONS AND FUTURE STUDIES

Implications and Problems Arising

The problems arising can be clearly seen: while the DML model serves as a comprehensive model for using multimedia and advanced learning technologies to achieve critical learning and develop lifelong learners, few students actually proceed beyond Stage 2 or 3—both the collaborative stages. The results falling from this are that so-called educationalists have their students afloat without winds in the doldrums. Students are using computers to chat and find (old foundational) knowledge, relating personal whims in Stage 1, and sharing interesting anecdotes in Stage 2, and not engaging academic knowledge-creation in Stage 3. While such depressing results were not seen in the present study, the obtained results were mixed. Some students could succeed to complete a full learning cycle of the model, and a couple went on to lifelong learning. But most undergraduate students found navigating the collaborative process of Stage 3 too difficult, despite the availability of scaffolding giving much additional structure to facilitate the required dialogue (from D- S+ Stage 2, to D+ S+ Stage 3).

It was also apparent that students found it difficult to move from Stage 3 to Stage 4. They reported that they could discover knowledge, views, and perspectives from other students and the World Wide Web and could make their own opinions from weighing these critically. However,

they reported difficulty in relating the theoretical perspective of Stage 3 to their own practical context in Stage 4 experientially. (A solution is given here next, rather than in the following section, for clarity.) Dialogue is very important in Stage 3, and a lot is needed. So, guided conversation is used. After largely achieving coconstruction of new understanding and knowledge, they need to move into Stage 4. To help them manage this, the tutor should increase dialogue even more by introducing synchronous conferencing. In the present study, personal presentations were made publicly to other students concerning the impact of the new knowledge on their lives and how they would try out new ideas in their lives. According to Moore (1993) the institution here should "take measures to reduce transactional distance by increasing the dialogue through use of teleconferencing" (p. 27). Students will lose some autonomy (A-) in going to synchronous mode, because they must become more empathic with others, but they will gain in dialogue (D+) and also in responsiveness to their own wants and needs and own context (with S- decrease in institutional structure).

Concerning the use of the World Wide Web and multimedia to promote students' learning, Herrington and Oliver (1999) reported that the higher-order thinking (of Stage 3 and Stage 4 here) was supported by using a situated-learning framework for relating the discussion to the student's own context. When using multimedia in situated learning, there was much less lower-order discussion and less social chat (Herrington & Oliver, 1999), indicating that multimedia could be applied to move students from Stage 1 to Stages 2 and 3. The implication here is that increasing the use of multimedia might have helped younger students cope better with the collaborative Stage 3.

Suggested Solutions

The various interactions between the student and tutor, student and other students, and student to content, and the quality, quantity, and frequencies of these constitute the academic dialogue in the educative process. The amount of dialogue needs to be carefully measured to suit each student's learning preferences and the task at hand. It is not true that simply increasing the amount of dialogue will solve all these interaction problems.

Adults generally need their prior experience and knowledge to be valued. Stage 3 entails collaborative argument. This needs an openness and receptiveness to have one's ideas be contradicted. Taking in new conflicting perspectives or information means, first, deconstruction of the existing cognitive knowledge network, where such deconstruction can be painful, especially when the prior understanding (like an old and trusted friend) has served the adult well to date. Concerning the uptake of learning technologies in their own courses, teachers, for example, have expressed a willingness to accept the innovation only insofar as it can be taken in small safe steps, permitting the teacher the safety-net option of recoursing to their proven methodology. The tutor should closely guide adults to moderate the amount of new conflicting information to prevent loss in their self-esteem. This is especially important during the Stage 3 collaborative argument and is not unimportant in the cooperative stages. Adults with a preference to field dependence (defined by Walter, 1998, as "those who gradually build towards generalisations about patterns from repeated exposure") will want to receive much information and likely enjoy cooperative learning in a group, while adults with a preference to field independence (defined by Walter, 1998, as "those who tend to see patterns and general principles in a flash of insight") are likely to want very much less input and may prefer the reflective process of collaborative learning in a group. The tutor is going to have a difficult time trying to moderate the amount of information proffered in student-to-student interactions.

To manage the quantity of new information, and the quality and frequency, the tutor could

direct cooperative massive exchanges away from Stage 2 and Stage 3 to a virtual "coffee-shop" set up expressly for this purpose, to keep the collaborative forum uncluttered. Then, the tutor will need to direct field-dependent learners to this virtual coffee-shop to assist their learning in their preferred way. This will also keep the online main forum clearer for the field-independent learners, during cooperative learning when field-dependent learners may be up-loading voluminous perspectives.

In Stage 3, there is a benefit to everyone to have diversity as wide as possible in different perspectives through which to test multiple hypotheses. Overloading to field-independent learners ought to be avoided, so careful use of hypertext is suggested here. For example, hypertext could be used in Stage 3 to give available links to reasons, keeping the main forum relatively uncluttered. It is necessary for the tutor to preascertain the field-dependence/independence of the participants and then closely guide each type separately, or to utilize some technique such as adaptive hypertext to accommodate these differences, equitably. The DML model indicates when, how, and why such adaptive hypertext will be useful. Using adaptive hypertext, the institution can provide extra interactivity to field-dependent students in the asynchronous collaborative stages, when there may be online silence among the field-independent learners. In the absence of adaptive hypermedia, the tutor should carefully tailor the amount of tutor-to-student messages to each type of learner.

Future Studies

Future studies are required to explore further why students find Stage 3 difficult to navigate through. In law and in health care and medicine, the collaborative critical-thinking skills in Stage 3 are especially important. Several institutions base their curricula now on trying to impart these skills using problem-based learning, though not all students prefer or choose this way of learning (for example, see Mangan, 1997, for law, and see Barrows, 1998, for medicine). Using this DML model, the current attention to problem-based learning processes can therefore be understood, and problem-based learning can be seen clearly in relation to the other ways of learning.

Since the present preliminary results were confounded by gender differences as well as by the use of English as a foreign language, further extended studies are underway. To investigate the correlation, if any, between the use of English as a foreign language and any potential overload (suggested by reduced reading and writing speeds by non-native-English users by Kawachi, 2002a, 2002b), identical courseware in various languages has been identified (namely Pocock & Richards, 1999), and students studying in their native language will be followed and comparatively measured.

Student motivations to learning online remain an area for further studies. How to initiate each and all the various intrinsic motivations to learn has been reported by Kawachi (2002c). However, further studies are warranted, because current taxonomies of adult motivations to learn pre-date multimedia learning technologies.

CONCLUSION

The DML model has been tested out in Japan, in large and small classes. Larger classes were divided into small groups of five or six students each. However, only in the small classes did students move successfully through the whole learning cycle. It was concluded that learning critical thinking using multimedia was better suited to graduate-level students or to small groups of tutor-guided undergraduate students. It was also concluded that tutor (the author) guidance was an important element and was too thinly spread while trying to manage five or six small groups simultaneously. In the larger classes, students did not achieve mastery of the collaborative learning,

despite trying to use the frameworks provided.

The deployment of learning technologies does not result simply in the status quo plus technology, but instead results in a new complex educational environment. The DML model tested out here provides a clear guide to technology users. However, the relative amount of time to be spent in each stage of this model is not prescribed and must be varied according to the students' own pace and according to the topic under study. Quality in learning outcomes can be defined as learning that has been achieved efficiently in terms of resources, is long lasting, and has personal meaning in the relevant desired context. To assure quality, the available learning technologies need to be utilized strategically. Different students naturally bring various learning preferences with them, and a single mode of teaching will be inappropriate. The advantage of multimedia is that multimedia can be designed to appeal to these various preferences simultaneously. The model of learning critical-thinking skills investigated here provides a scaffold to all the agents involved in education, including the administrators, teachers, nonacademic support, and students. This model as a scaffold serves as a cue and support to all these agents. Some students (or some teachers) might be uncomfortable in a particular stage of this model, where learning proceeds through a nonpreferred way. For example, field-dependent learners may prefer the synchronous cooperative stages, while field-independent learners may prefer the asynchronous collaborative stages. Nevertheless, critical thinking is a universally avowed desirable goal in adult education, and adults need to acquire these skills and should be strategically flexible in their approaches to study. The model dictates when switching to another approach is required, to proceed optimally and learn efficiently the full repertoire of skills that interlink for critical thinking. There is no argument that some courses may utilize only one way of teaching and learning. This model shows how such courseware might be improved for all-round human resource development.

Computer-mediated communications are being utilized for an increasing number of students in both conventional classrooms and at a distance, in synchronous mode and in asynchronous mode. This could suggest that more research and design resources might be forthcoming. Yet there is little research to date on why and when to utilize these technologies. In some studies, synchronous videoconferencing technologies have been bought and can technically connect the various agents for the learning process, but the tutors and institutions aim for collaborative learning, for which the synchronous technology is inappropriate and unsuccessful. Remarkable efforts are being made by many institutions worldwide to apply these new technologies for learning, yet the instructional design and the technology selected continue to be important factors causing the failures to achieve higher-order thinking skills (Abrami & Bures, 1996, p. 37). The present DML model is the only practical model proposed to date for selecting and ordering the utilization of learning technologies for acquiring critical-thinking skills. As such, the DML model constitutes an intelligent syllabus to be tested out further.

REFERENCES

Abrami, P. C., & Bures, E. M. (1996). Computer-supported collaborative learning and distance education. *American Journal of Distance Education, 10*, 37–42.

Barrows, H. (1998). *Problem based learning.* Southern Illinois University School of Medicine. Retrieved January 10, 1999 from the World Wide Web: http://edaff.siumed/dept/index.htm

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1997). *Women's ways of knowing: The development of self, voice and mind* (10th anniversary ed.). New York: Basic Books.

Bork, A. (2001). What is needed for effective learning on the Internet. Special issue on curriculum, instruction, learning and the Internet. *Educational Technology and Society*, (in press). Retrieved June 10, 2002 from the World Wide Web: http://www.ics.uci.edu/~bork/effectivelearning.htm

Boud, D. (1988). Moving toward student autonomy. In D. Boud (Ed.), *Developing student autonomy in learning* (2nd ed.) (pp. 17–39). London: Kogan Page.

Briguglio, C. (2000). Self directed learning is fine—If you know the destination! In A. Herrmann, & M. M. Kulski (Eds.), *Flexible futures in tertiary teaching—Proceedings of the 9th Annual Teaching Learning forum*, February 2–4, 2000, Curtin University of Technology, Perth, Australia. Retrieved May 14, 2003 from the World Wide Web: http://cea.curtin.edu.au/tlf/tlf2000/briguglio.html

Brookfield, S. D. (1987). *Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting.* San Francisco, CA: Jossey-Bass.

Bynner, J. M., & Romney, D. M. (1986). Intelligence, fact or artefact: Alternative structures for cognitive abilities. *British Journal of Educational Psychology, 56,* 13–23.

Chen, Y. -J., & Willits, F. K. (1999). Dimensions of educational transactions in a videoconferencing learning environment. *American Journal of Distance Education, 13*(1), 45–59.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process.* Lexington, MA: D.C. Heath and Company.

Gibbs, G., Morgan, A., & Taylor, E. (1984). The world of the learner. In F. Marton, D. Hounsell, & N. J. Entwistle (Eds.), *The experience of learning* (pp. 165–188). Edinburgh: Scottish Academic Press.

Gunawardena, C., Plass, J., & Salisbury, M. (2001). Do we really need an online discussion group? In D. Murphy, R. Walker, & G. Webb (Eds.), *Online learning and teaching with technology: Case studies, experience and practice* (pp. 36–43). London: Kogan Page.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 397–431.

Halpern, D. F. (1984). *Thought and knowledge: An introduction to critical thinking.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Harrison, C. (2002). Hypertext links: Whither thou goest, and why. *First Monday, 7*(10). Retrieved October 10, 2002 from the World Wide Web: http://firstmonday.org/issues/issue7_10/

Herrington, J., & Oliver, R. (1999). Using situated learning and multimedia to investigate higher-order thinking. *Journal of Educational Multimedia and Hypermedia, 8*(4), 401–422. Retrieved May 6, 2003 from the World Wide Web: http://dl.aace.org/9172

Holmberg, B. (1983). Guided didactic conversation in distance education. In D. Sewart, D. Keegan, & B. Holmberg (Eds.), *Distance education: International perspectives* (pp. 114–122). London: Croom Helm.

Kaplan, H. (1997). Interactive multimedia & the World Wide Web. *Educom Review, 32*(1). Retrieved May 21, 2003 from the World Wide Web: http://www.educom.edu/web/pubs/review/reviewArticles/32148.html

Kawachi, P. (2000). *Why the sun doesn't rise: The impact of language on the participation of Japanese students in global online education*. Unpublished MA ODE Thesis, Open University,

Milton Keynes, UK. Available from the author by e-mail: kawachi@kurume-shinai.ac.jp

Kawachi, P. (2002a). Poverty and access: The impact of language on online collaborative learning for Japanese learners. In H. P. Dikshit, S. Garg, S. Panda, & Vijayshri (Eds.), *Access & equity: Challenges for open and distance learning* (pp. 159–170). New Delhi: Kogan Page.

Kawachi, P. (2002b). On-line and off-line reading English rates: Differences according to native-language L1, gender, and age. *Proceedings of the 16th annual conference of the Asian Association of Open Universities*, Seoul, Korea, November 5–7. Retrieved January 10, 2003 from the World Wide Web: http://www.aaou.or.kr

Kawachi, P. (2002c). How to initiate intrinsic motivation in the on-line student in theory and practice. In V. Phillips et al. (Eds.), *Motivating and retaining adult learners online* (pp. 46–61). Essex Junction, VT: Virtual University Gazette. Retrieved August 25, 2002 from the World Wide Web: http://www.geteducated.com/vug/aug02/Journal/MotivateRetain02.PDF

Kawachi, P. (2003a). Vicarious interaction and the achieved quality of learning. *International Journal on E-Learning, 2*(4), 39-45. Retrieved January 16, 2004 from the World Wide Web: http://dl.aace.org/14193

Kawachi, P. (2003b). Choosing the appropriate media to support the learning process. *Journal of Educational Technology, 14*(1&2), 1–18.

Kawachi, P. (2003c). Initiating intrinsic motivation in online education: Review of the current state of the art. *Interactive Learning Environments, 11*(1), 59-81.

Laurillard, D. (2002). *Rethinking university teaching* (2nd ed.)*: A conversational framework for the effective use of learning technologies.* London: RoutledgeFalmer.

Lyons, R. E., Kysilka, M. L., & Pawlas, G. E. (1999). *The adjunct professor's guide to success: Surviving and thriving in the college classroom.* Needham Heights, MA: Allyn & Bacon.

Mangan, K. S. (1997). Lani Guinier starts campaign to curb use of the Socratic method. *Chronicle of Higher Education*, (11 April), A12–14.

McKinnon, J. W. (1976). The college student and formal operations. In J. W. Renner, D. G. Stafford, A. E. Lawson, J. W. McKinnon, F. E. Friot, & D. H. Kellogg (Eds.), *Research, teaching, and learning with the Piaget model* (pp. 110–129). Norman, OK: Oklahoma University Press.

McLoughlin, C., & Marshall, L. (2000). *Scaffolding: A model for learner support in an online teaching environment.* Retrieved May 14, 2003 from the World Wide Web: http://cea.curtin.edu.au/tlf/tlf2000/mcloughlin2.html

Moore, M. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). London: Routledge.

Neubauer, M. (2003a). *Asynchronous, synchronous, and F2F interaction.* Online posting May 23 to the Distance Education Online Symposium. Retrieved May 23, 2003 from the World Wide Web: http://lists.psu.edu/archives/deos-l.html

Neubauer, M. (2003b). *Number of online participants.* Online posting January 22nd to the Distance Education Online Symposium. Retrieved January 22, 2003 from the World Wide Web: http://lists.psu.edu/archives/deos-1.html

Palincsar, A. S. (1986). The role of dialogue in providing scaffolding instruction. *Educational Psychologist, 21,* 73–98.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme.* New York: Holt, Rinehart and Winston.

Piaget, J. (1977). Intellectual evolution from adolescence to adulthood. In P. N. Johnson-Laird, & P. C. Wason (Eds.), *Thinking: Readings in cognitive science*. Cambridge, UK: Cambridge University Press.

Pocock, G., & Richards, C. D. (1999). *Human physiology: The basis of medicine*. Oxford : Oxford University Press (and same courseware in Japanese, in French, and in Spanish).

Probst, R. E. (1987). Transactional theory in the teaching of literature. *ERIC Digest* ED 284 274. Retrieved April 24, 2002 from the World Wide Web: http://www.ed.gov/databases/ERIC_Digests/ed284274.html

Renner, J. S. (1976a). Formal operational thought and its identification. In J. W. Renner, D. G. Stafford, A. E. Lawson, J. W. McKinnon, F. E. Friot, & D. H. Kellogg (Eds.), *Research, teaching, and learning with the Piaget model* (pp. 64–78). Norman, OK: Oklahoma University Press.

Renner, J. S. (1976b). What this research says to schools. In J. W. Renner, D. G. Stafford, A. E. Lawson, J. W. McKinnon, F. E. Friot, & D. H. Kellogg (Eds.), *Research, teaching, and learning with the Piaget model* (pp. 174–191). Norman, OK: Oklahoma University Press.

Reynolds, B. (2003). Synchronous instruction in D/E. Online posting May 27 to the Distance Education Online Symposium. Retrieved May 27, 2003 from the World Wide Web: http://lists.psu.edu/archives/deos-l.html

Rosenshine, B., & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership, 49*(7), 26–33.

Tennant, M.C., & Pogson, P. (1995). *Learning and change in the adult years: A developmental perspective*. San Francisco, CA: Jossey-Bass.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Walter, C. (1998). Learner independence: Why, what, where, how, who? *Independence*: Newsletter of the IATEFL Learner Independence Special Interest Group, *21*, 11–16.

Wang, H. (2002). The use of WebBoard in asynchronous learning. *Learning Technology Newsletter, 4*(2), 2–3. Retrieved June 10, 2002 from the World Wide Web: http://lttf.ieee.org/learn_tech/

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89–100.

Zimmer, B. (1995). The empathy templates: A way to support collaborative learning. In F. Lockwood (Ed.), *Open and distance learning today* (pp. 139–150). London: Routledge.

This work was previously published in Interactive Multimedia in Education and Training, edited by S. Mishra and R.C. Sharma, pp. 1-28, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.12
Web-Based Multimedia Children's Art Cultivation

Hao-Tung Lin
National Chi-Nan University, Taiwan, R.O.C.

Herng-Yow Chen
National Chi-Nan University, Taiwan, R.O.C.

INTRODUCTION

With the rapid advance in Web and multimedia technologies, authoring various types of multimedia content and distributing them on the Web has been very popular for many years. These technologies are applied in e-learning extensively, such as from language learning (e.g., ESL) to professional domain knowledge (e.g., computer science). In contrast, e-learning systems focusing on art domains, especially for kids or teenagers, are few. This is a notable shortcoming, because from a technical viewpoint, current advances in multimedia technology via the Web promise this kind of application. On the other hand, compared with technologies needed for more general-purpose knowledge, cultivating children's art through e-learning technology needs much more edutainment ingredients – it must be interesting and interactive and offer multimedia. Realizing this kind of e-learning is really a challenge, not only from a pedagogical viewpoint (the first ingredient) but also technical ones (the latter two ingredients).

In this article, we describe how our framework design for online authoring and presentation works. The goal of this framework is to provide a universal platform that enables students to learn more actively through sharing their own pieces easily with other learners. Peers and teachers can comment on students' work for further discussion or instruction.

To this end, a multimedia authoring and presentation tool named "My E-card" (http://media.csie.ncnu.edu.tw/haotung/myecard/) has been designed to allow students to combine different-media objects (such as a painting object, typing object and music object) into a time-ordered, synchronized multimedia document (i.e., animated sound painting). Students can import any existing media objects (e.g., image files or MIDI files) in cyberspace through a Universal Resource

Locator (URL), or create new ones from different supporting tools, such as static painting, writing an essay or composing music. We use the XML format to describe the multimedia objects and their temporal, spatial relationship metadata because of XML's high extensibility and flexibility (W3C, 2004; Villard, Roisin, & Layada, 2000). Students can resume their work at other places. They don't have to worry about data integrity or the presentation consistency of the unfinished work deposited in the server. At any stage, current piecework can be played out with synchronization to preview the result.

Research has indicated that both competence and confidence are keys to the success of active learning (Koutra, Kastis, Neofotistos, Starlab, & Panayi, 2000; Jeremy, Roy, Christopher, Douglas, & Barbara, 2000). Experimental results show that our present work enforces the highly interactive creation process, which involves acts of media creating and further authoring – an approach that leads to personal competence. Moreover, playing composite multimedia work with a synchronized manner and sharing the great work with friends reinforce personal confidence.

SYSTEM FRAMEWORK

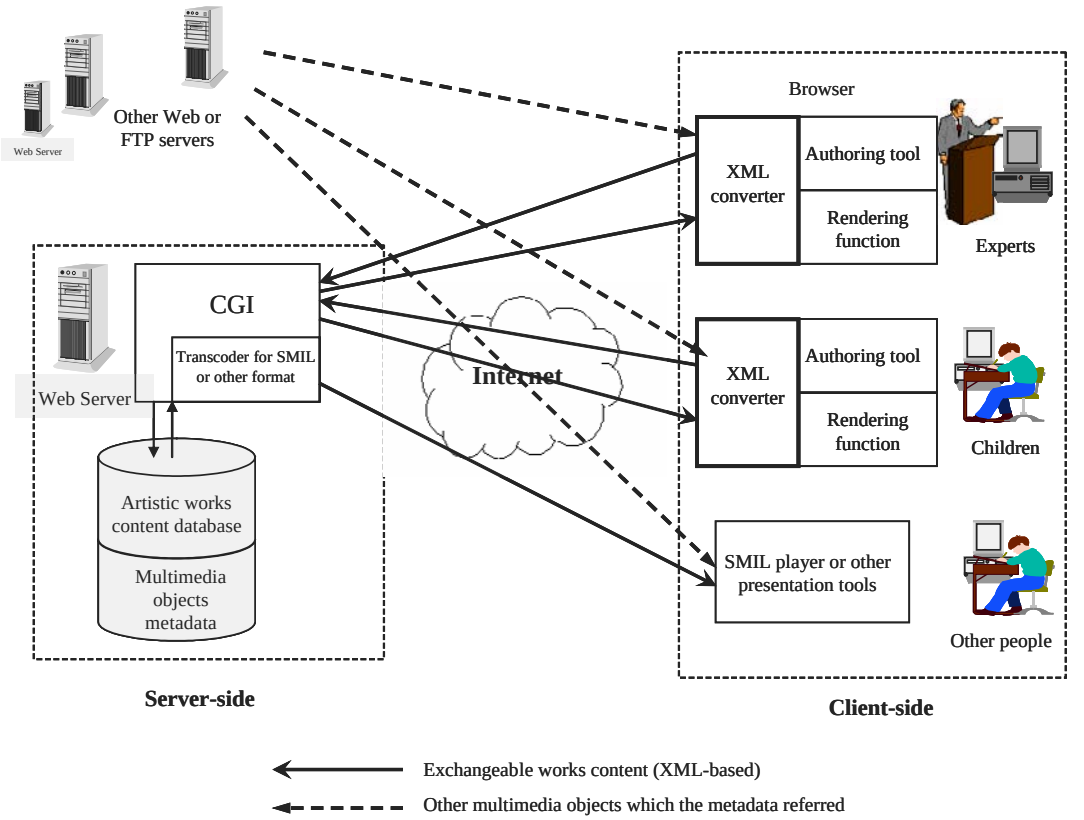
Figure 1 illustrates the proposed framework over existing web architecture, which is basically a client-server architecture: clients for authoring and presentation, servers for metadata storage and format exchange. The customized multimedia authoring and presentation program will be downloaded from the server into the client's browser and automatically executed. The program should provide users (e.g., students, teachers, experts and others) with most friendly multimedia-authoring functions and presentation experiences. All the authoring results will be transmitted to the server for storage and sharing. The server-side application gateway should maintain the meta-information of the composed, submitted artwork and handle the content exchange necessary for a different presentation tool (e.g., SMIL player). The XML-based metadata describing how different sources of media objects are composed mainly consists of three types of information: the URL of media objects (where to locate them), temporal information (when to display them) and spatial information (where to place them). The rendering function of the presentation program needs those metadata to make a best performance to users. With the help of Web technology, the framework can easily reuse a variety of existing multimedia object resources (such as images, video and audio URLs) which need not be stored in the server.

ENABLE CHILDREN TO BE INTERESTED IN THE SYSTEM

The environment and tools supported by the system may help children to cultivate their art capacity. However, how to attract children to use them actively is an important topic: It decides whether the system is worthy. We apply the following guidelines that can attract children to join this site actively.

1. Children's works should be able to be shared and viewed easily.Because it is an environment where everyone can learn through viewing other people's work, the Web site exhibits all the works created by children. Children can send their work as an e-card to relatives and friends. Through this Web-based system, people can see their works anywhere, anytime. (Bandura, 1986)
2. Children can get feedback about their work from others.Art appreciation is sometimes subjective. Many people may have different answers to the question "Is this piece a good work?" Everyone wishes to get feedback from other people, especially positive com-

Figure 1. The Web-based multimedia authoring and presentation framework



ments. For this reason, some experts in the arts are invited as online reviewers who, from time to time, examine online works and comment on them through Web interface or e-mail. (Skinner, 1968; Bandura, 1986).

3. An online contest with voting encourages children to perform.We encourage registered students to take part in an e-card contest on the basis of some specific topics. In addition to reviewers, Web visitors can vote for the work they think the best. Children must hope that their work can be a popular one. This drives contestants to keep improving their own piece. (Sulzer-Azaroff, & Mayer, 1986; Bandura, 1986)

IMPLEMENTATION

We use the Apache (Apache, 2004) as our HTTP server, incorporated with the PHP language as the server-slide Common Gateway Interface (CGI). A large number of Web users can view Flash-based Web content in their browsers (OpenSWF.org, 2004). The Macromedia (Macromedia Flash, 2004) Flash technology has been a de facto platform because of its highly interactive capability and multimedia (such as gif, jpg, wav and mp3) format support. To provide as friendly as possible an interface for end users, the Flash technology incorporated with the JavaScript Dynamic HTML control (DHTML, 2004; Dynamic Drive, 2004)

are used to develop the authoring and presentation program.

Figure 2 shows some examples of the operation in My E-card, the major authoring tool in this system. Figure 3 illustrates how the rendering function presents the children's work. There are two major media in this presentation – visual appearance and audio (music). The music can be played in several modes: background music, repeated until the end of presentation; introduction music, played just in the beginning; or throughout the entire presentation. In the third mode, the animated painting actions should be finished by the end of the music, so the rendering function normalizes all the timestamps of the significant painting actions and fits them into the time scale of the music. The painting objects appear one by one according to the time stamps, as shown in Figure 4.

The presentations of the art works are very vivid and interesting. They reflect the painting process step by step and with sound effects. The scenario of the presentation looks like a painter who is painting on the spot, accompanying the music playback. We got a lot of positive feedback from online visitors since the Web site was announced in 2003. Those people created their art works through the proposed tools and operated them well. The quality of the works they made is quite good, and sometimes beyond our expectations. Some adults even told us that they think the tools we proposed are very funny and are willing to visit the site more frequently.

Using the XML-based format to represent the metadata makes the integration of different types of media in different presentation platforms much easier. The cost of storing a large number of multimedia objects can be decreased, and the multimedia objects created by different authors can also be reused easily.

Figure 2. Some examples of the operation in My E-card. (a) The main interface of My E-card. (b) The ready-objects chooser. Children can choose the ready objects to compose a scene. (c) The function of music chooser and synchronization configuration. (d) The animated presentation process in the presentation interface. A slide bar on the top of the view indicates the presentation time. A virtual moving pen over the currently rendered object is for reality.

(a) (b)

(c) (d)

Figure 3. Logical view of the development rendering function

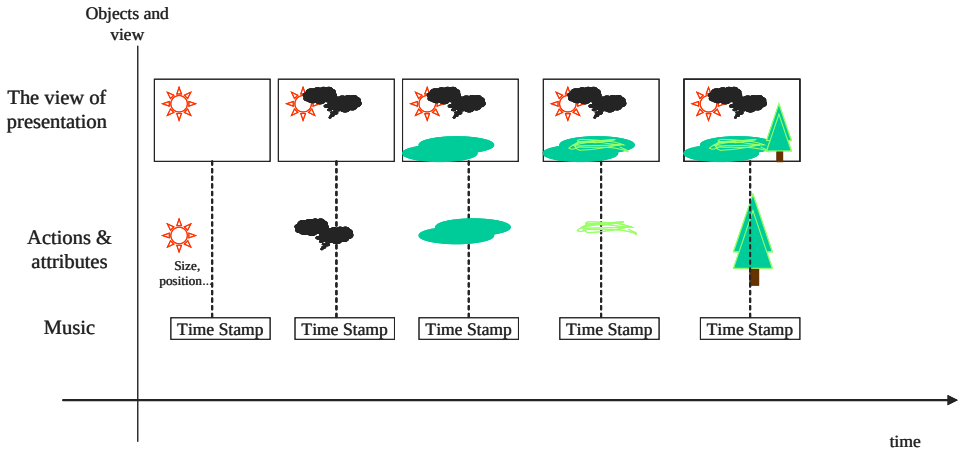
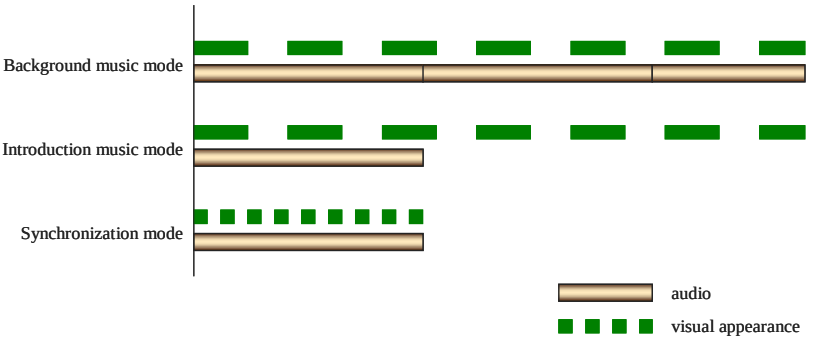


Figure 4. The diagram of the presentation modes



FUTURE TRENDS

In our current system, we have supported some authoring tools and the player. The works children made are stored in XML format. The XML format is easy to distribute, share and reuse, so our authoring tools and players can integrate and present the different types of the works easily. The definition of the XML tag we used has been designed for our tools. If the data can be converted into SMIL format or other standards, the works would be distributed and shared with others more easily (Boll, Klas, & Wandel, 1999; Martin, & Mulhem, 2000).

On the basis of the proposed framework, we can develop more interesting authoring and production tools for different types of art learning and cultivation, such as music composing. The tool generates MIDI data, saved in an XML-based format. Our authoring tool, My E-card, can import the finished product (e.g., MIDI file). Then, children can touch more types of art.

CONCLUSION

In this article, we described a Web-based learning environment for cultivating children's art capacity. A prototype has been implemented on a Web site especially designed for young student art learning, which has been conducted by the Military of Education, Taiwan, since 2003. The proposed Web-based multimedia authoring and presentation framework can facilitate online

artwork creation, sharing and reuse in e-learning applications. Through the easy-to-use interface of the customized tools, students can create their own pieces and share with friends. To make the presentation more interesting, a synchronization module in rendering function replays a vivid, animated multimedia presentation in a form as close as possible to the original.

To advance the worth of the framework and the system, we are inviting artists and professionals in different art or education domains to assist us in designing an objective evaluation mechanism. The results of evaluation will offer us more aspects of system promotion.

REFERENCES

Apache Software Foundation, The. Retrieved August 3, 2004 from *www.apache.org/*

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs: Prentice-Hall.

Boll, S., Klas, W., & Wandel, J. (1999). A cross-media adaptation strategy for multimedia presentations. *ACM Multimdedia '00 Proceedings*, 37-46.

DHTML Lab. Dynamic HTML tutorials, DHTML scripts, programming, tools, code, and examples. Retrieved August 3, 2004, from *www.web reference.com/dhtml/*

Dynamic Drive DHTML & JavaScript code library. Retrieved August 3, 2004, from *http://dynamicdrive.com/*

Koutra, C., Kastis, N., Neofotistos, G., Startlab, & Panayi, M. (2000). Interactive learning environments for children: User interface requirements for a Magic Mirror and Diary Composer Environment. *Proceedings of One-Day Workshop on Interactive Learning Environments for Children*, Athens, Greece, March 1-3.

Macromedia Flash Developer Center. Retrieved August 3, 2004 from *www.macromedia.com/devnet/mx/flash/*

Martin, H., & Mulhem, P. (2000, July). A comparison of XML and SMIL for on the fly generation of multimedia documents from databases. *SCI Conference 2000, Proceedings of the 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2000), 12, Computer Science and Engineering: Part I, July 2000*, Orlando, Florida, 11-16.

OpenSWF.org. The source for Flash File format information. Retrieved August 3, 2004, from *www.openswf.org/*

PHP: Hypertext Preprocessor. Retrieved August 3, 2004, from *www.php.net/*

Roschelle, J.M., Pea, R.D., Hoadley, C.M., Gordin, D.N., Means, B.M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children and Computer Technology, 10*(2), Fall/Winter.

Skinner, B.F. (1968). *The technology of teaching.* New York: Appleton-Centry-Corfts, Prentice Hall Div.

Sulzer-Azaroff, B., & Mayer, G. (1986). *Achieving education excellence using behavioral strategies.* New York: Holt, Rinehart & Winston.

Villard, L., Roisin, C., & Layada, N. (2000). An XML-based multimedia document processing model for content adaptation. *Digital Documents and Electronic Publishing (DDEP00), LNCS*, Springer Verlag.

W3C. World Wide Web Consortium. *Extensible Markup Language (XML).* Retrieved 2004, from *www.w3.org/XML/*

KEY TERMS

E-Learning: Education via the Internet, network, or standalone computer. Network-enabled transfer of skills and knowledge. e-Learning refers to using electronic applications and process to learn. e-Learning applications and processes include Web-based learning, computer-based learning, virtual classrooms, and digital collaboration. Content is delivered via the Internet, intranet/extranet, audio or video tape, satellite TV, and CD-ROM.

SMIL: Synchronized Media Integration Language, a markup language designed to present multiple media files together. For instance, instead of using a video with an integrated soundtrack, a separate video and sound file can be used and synchronized via SMIL. This allows users to choose difference combinations, e.g., to get a different language soundtrack, and permits text transcripts to be optionally presented; both options have accessibility benefits.

Synchronized Multimedia Document: Multimedia systems usually integrate different types of data streams, including continuous media (e.g., audio and video) and discrete media (e.g., text and still images). Media data must be presented at precise time instants defined by the rate of presentation. A media data stream schedules presentation of samples within a given time base. In this way, objects in the same time base are synchronized.

This work was previously published in Encyclopedia of Distance Learning, Vol. 4, edited by C. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers, and G.A. Berg, pp. 2004-2008, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.13
Student-Generated Multimedia

Mathew Mitchell
University of San Francisco, USA

INTRODUCTION

This entry looks at the role of student-generated multimedia (SGM) in helping students more effectively achieve meaningful outcomes. The entry first looks at the theory and research behind multimedia learning and then goes on to address the specific case of student-generated multimedia. Mayer (2001) defined multimedia as the presentation of material using both words and images, and then subsequently defined multimedia instruction as a "presentation involving words and pictures that is intended to foster learning" (p. 3). The implications of these definitions are important because they delineate two key aspects to thinking about multimedia. First, multimedia products do not need to use video, animation, or interactivity. More importantly, Mayer's definitions focus on multimedia's potential benefits as a learning tool rather than as a technological device.

While there has been a strong body of research supporting the learning benefits of multimedia under specific conditions (e.g., see Mayer, 2001), there has been little research done on the potential benefits of student-generated multimedia. Within the larger context of educational thinking, this seems odd as there is general support for the basic idea that students learn better if they are the authors or creators of significant learning products (e.g., see Reigeluth, 1999). The theme of the constructivist research base has been that students tend to construct a deeper understanding of content if challenged to solve ill-defined problems within a relevant, real-world context. More specifically, there has been a consistent body of research that indicates students learn conceptually demanding material better when they construct self-explanations (Chi, 2000). One of the hallmarks of well-designed student-generated multimedia challenges is that they seem to invoke the self-explanation effect. Despite these factors, relatively little attention has been paid to the potential benefits of SGM. In a recent book (Brown, 2000) containing 93 vignettes from America's "most wired campuses," only five of the vignettes were about student-generated products. A more recent article investigating computer-using activities of both teachers and students (Zhao & Frank, 2004) used no measure of student-generated products. They looked at high-school classrooms, but the nearest measure to SGM was labeled "student inquiry," and their

results indicated that less than 14% of students engaged in these inquiry kinds of activities such as conducting student research using the Web. Given that computer technology provides great opportunities for students to relatively easily and effectively create significant learning products, it would seem reasonable that future research into computer-learning technologies explore the arena of SGM.

WHY MULTIMEDIA?

If the learning of specific material is easily accomplished by most students, then there is no need to incorporate multimedia instruction. For example, creating a multimedia product is typically more time intensive relative to the creation of a new text product. Thus, using multimedia makes the most sense if the specific learning challenge for students is difficult for them to master using traditional methods of instruction. In such a case, multimedia products may provide the extra instructional boost that will make a difference to learners mastering difficult material. Mayer's (2001) cognitive theory of multimedia provides an explanation for *why* the appropriate use of multimedia can be especially helpful in learning complex concepts.

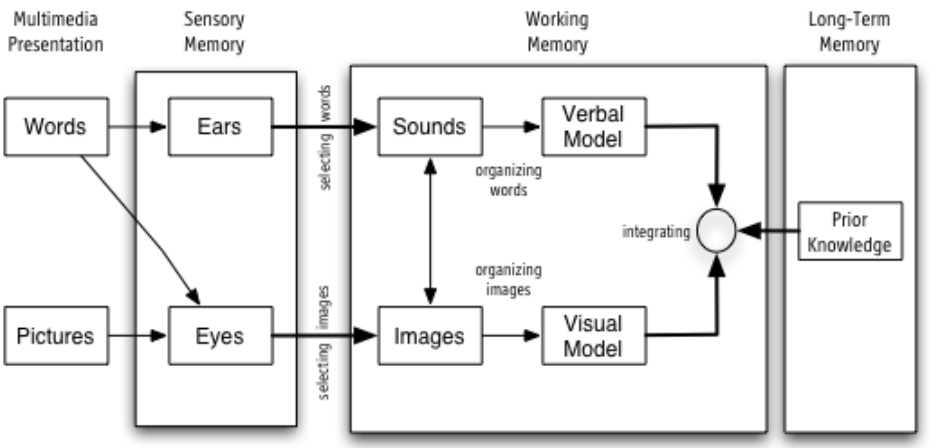
Consistent with the above line of thinking, SGM is probably best incorporated into the learning environment when the conceptual material under study is both difficult and essential for students to master. After all, to create multimedia, students need to learn about creating images and audio, they still need to know how to use text, and typically the most difficult design challenge is integrating all of these elements into one cohesive product. However, multimedia learning challenges may more effectively facilitate students' learning of complex material at a much deeper level than possible with exams or papers. For example, take the case of analysis of variance for doctoral students. This statistical concept and technique is both challenging for them to understand well and essential for their future work as academics. Under these conditions it may make sense to incorporate SGM learning challenges into the classroom. However, for simpler conceptual material such as finding the mean, it is likely an inappropriate use of time and resources to use SGM challenges. Of course, what is challenging and essential for students is dependent on grade level and a variety of other factors. Yet, given that a teacher, a department, or a school has identified particularly difficult-to-achieve outcomes, they may want to consider using SGM learning challenges. An additional benefit to using a multimedia approach to learning conceptually challenging material is that it is much more flexible compared to text-based approaches (such as essay papers) in terms of meeting the needs of learners with a wide array of learning styles. This is because multimedia allows learners to demonstrate their understanding using more than one mode of communication. As this entry will discuss in more detail below, the main potential benefits of well-designed SGM instructional challenges is that they tend to facilitate active learning in students through self-explanation and relevance.

A COGNITIVE THEORY OF MULTIMEDIA LEARNING

It is difficult to understand the value of SGM without a grounding regarding the cognitive benefits of multimedia learning in general. A leading cognitive researcher in terms of the learning benefits of multimedia is Richard Mayer (2001). While other researchers have proposed models of multimedia learning (e.g., Hede, 2002), Mayer's model (see Figure 1) seems the most complete based on a strong foundation of closely aligned research.

The focus of Mayer's (2001) model is on meaningful learning (operationally measured as transfer or problem solving in most studies).

Figure 1. Mayer's cognitive theory of multimedia learning



However, if the type of learning outcomes desired can be described as the retention of facts, then multimedia formats for learning seem no better than other formats such as text (Mayer).

Mayer's (2001) model describes what optimally happens when active learning is going on. Specifically, Mayer defines active learning as the three-pronged process of selecting, organizing, and integrating information. The basic idea is that when active learning is optimized, then meaningful outcomes can be more effectively reached. In his model, information is presented through words or pictures. Note that words can be taken in primarily through the ears or the eyes depending on whether one is using audio narration or written text. Mayer's model predicts that we select from the presented information a subset of it that will be processed in working memory. This selection of information (in the form of sounds and images) is next organized into coherent models (both verbal and visual). Finally, in the last stage of active learning, integration occurs between the verbal model, the visual model, and the learner's prior knowledge. Mayer's thinking is that when a learner has both a verbal and visual model of the information, as well as some prior knowledge that can be used to make sense of the information, then the process of integration and subsequent understanding of the material will be enhanced.

One key is how to prime active learning so that meaningful outcomes such as problem solving can be met. In essence, Mayer (2001) proposed that there are three major cognitive priming agents for active learning: cognitive load, social agency, and self-explanation. The vast majority of existing research has been on cognitive load. The idea behind cognitive load is that our visual and auditory capacities are extremely limited. Since these channels can easily become overloaded, one potential benefit of multimedia learning is that it can present complex information in a format that minimizes potential overloads. How does one do this? Mayer and Moreno (2003) proposed that there are five specific overload situations, and they put forth nine solutions to those problems (based on a review of the current research). Those nine solutions were called off-loading, segmenting, pre-training, weeding, signaling, aligning words and pictures, eliminating redundancy, synchronizing, and individualizing. Many of the nine solutions can be optimally met using multimedia solutions that integrate audio narration and visual images while reducing or eliminating the use of visual text. Mayer's simplest guiding principle for creating multimedia is to construct a concise narrated

animation, a type of multimedia presentation that uses synchronized audio narration along with visual material that is concise and uses a meaningful structure (such as a cause-and-effect chain).

Of the nine solutions, the one with the least amount of research behind it is individualizing. For example, students who have low prior knowledge but high spatial ability best learn from multimedia presentations. However, what about other types of learners? How can they best be accommodated? This is a rich area for future research (e.g., see Mayer and Massa, 2003). Cognitive load may also be age related. For instance, Mann, Newhouse, Pagram, Campbell, and Schulz (2002) provided recent evidence that some forms of cognitive overload may not be as critical for younger learners (12 years old in their study). On the other hand, Van Gerven, Paas, Van Merrienboer, Hendriks, and Schmidt (2003) provide evidence that multimedia instruction may be more effective for elderly learners (64.5 years was the mean age of their sample) due to a reduction in perceived cognitive load.

While cognitive load has received the majority of the research attention, the other two cognitive priming agents are equally fascinating and deserve further research. The priming agent called social agency refers to the idea that social cues can prime the conversation schema. The benefit of this is that the learner will expend greater cognitive effort when they feel they are part of a conversation than when they are listening to a more formal narration. Two recent studies (Mayer, Fennell, Farmer, & Campbell, 2004; Moreno & Mayer, 2000) found that simply using a script with "I" and "you" to better connect with the audience had better results in terms of student learning than using the standard third-person narrative approach. One can surmise there may be additional ways to prime the conversation schema.

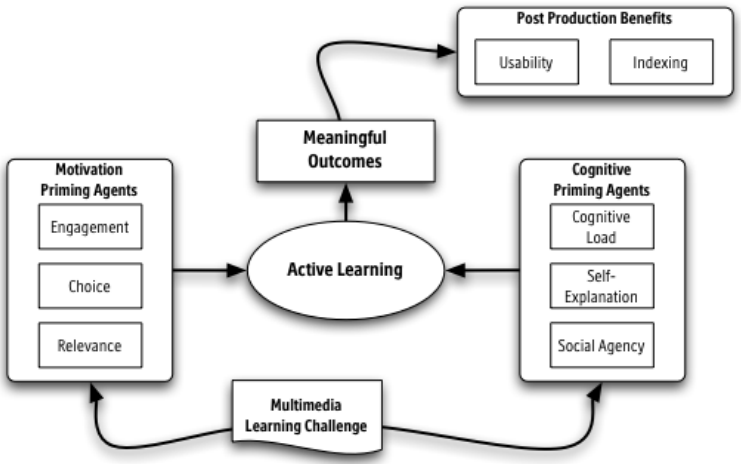
The final priming agent called self-explanation refers to the idea that if a learning experience can get the learner to be involved in self-explanation, then that will result in greater cognitive engagement. For instance, Mayer, Dow, and Mayer (2003) tested the idea that if the learner is encouraged to answer conceptual questions during the learning experience, then that will result in a higher level of learning. Indeed, they found support for their conjecture through the simple use of presenting learners with a conceptually demanding question that the learner could not fully answer before receiving the multimedia presentation. Their explanation of the results was that such a rich prequestion primes the learner to be more actively engaged in learning the material (in order to answer the prequestion) while watching the multimedia presentation.

TOWARD A THEORY OF STUDENT-GENERATED MULTIMEDIA

Mayer (2001) has presented the most compelling model of multimedia learning to date. While Mayer's model is likely incomplete, it nonetheless offers a foundation for understanding when and how student-generated multimedia may be effective. Using Mayer's model as a guide, a model for SGM is suggested by the author and presented in Figure 2.

The first focus of this proposed model is on meaningful outcomes. In the general model of multimedia learning, outcomes were operationally defined as learning measured by transfer tasks or problem solving. Similarly, SGM is probably best used when the conceptual material is both difficult for students and essential for them to master such as the analysis of variance for doctoral students in a statistics course. The reason for pragmatically limiting SGM to material that is difficult and essential has to do with the time-intensive nature of such projects. It is simply not worth an instructor's or student's time to invest in SGM unless it facilitates the accomplishment of difficult and essential outcomes.

Figure 2. Student-generated multimedia model



The proposed model of SGM posits that the most expeditious way to have students reach those meaningful outcomes is through active learning. The purpose of using an SGM learning challenge is to encourage key priming agents to be activated. This model includes Mayer's (2001) three key cognitive priming agents: cognitive load, social agency, and self-explanation. In addition the model posits that there are three motivational priming agents: engagement, choice, and relevance.

Research suggests that student-generated products in general can be motivating to students if the constraints for the products fit key criterion. One key area of supportive research comes from Benware and Deci (1984), who put forth their own hypothesis for active learning. Specifically, they defined active learning as learning in order to teach others. They hypothesized that active learning where the student is "... learning material to teach it will lead to enhanced learning and to a more positive emotional tone than learning material to be tested on it, even when the amount of exposure to the material being learned is the same" (p. 756). Benware and Deci's study, as well as some subsequent studies, provide general support for this approach. Their hypothesis could be framed as a specific instantiation of a constructivist learning environment and as supportive of the self-explanation aspect of Mayer's model. Importantly, their research suggested that not only did students learn better when provided with an active-learning prompt, but that they also felt more engaged with learning the material. Other motivation research (e.g., Mitchell, 1993; Schraw and Lehman, 2001) indicates that there are key motivational priming agents activated when a learning environment is high in situational interest (i.e., an interest generated due to the environment rather than a preexisting individual interest). Key aspects of high-situational-interest learning environments include engagement, choice, and relevance (or meaningfulness).

While many of the motivational and cognitive priming agents for active learning might be incorporated into a nonmultimedia student-generated product, multimedia pragmatically offers the simplest way to integrate all of the key cognitive and motivational priming agents for active learning. In addition, SGM uniquely provides the postproduction benefits of usability (the resulting product can be shared and viewed by current and future students) and indexing through the creation of a multimedia library.

Cognitive Agents

The SGM model suggests instructional practices based on the cognitive priming agents of cognitive load, social agency, and self-explanation. Cognitive-load research has generated the most recommendations. At a very basic level, this research suggests that SGM both off-load (i.e., reduce cognitive load by using more than one processing channel) and synchronize (i.e., reduce cognitive load by synchronizing visual and auditory material) when creating a product (Mayer & Moreno, 2003). These are extremely basic suggestions as they should be a part of any multimedia presentation. At a deeper level, cognitive-load research also suggests that students make sure to provide a meaningful structure to their products (Mayer, 2001). Two approaches that can help toward this end are conciseness (i.e., exclude extraneous material and highlight essential material) and organization (providing an appropriate structure). The most ambiguous level of suggestions for cognitive load surround the need for interactivity to meet individual difference needs. One interactivity approach that can help is self-pacing, which reduces cognitive load by allowing the learner to control the pace of the presentation (Mayer, 2001). However, it is likely there are additional techniques that would also meet individual difference needs that have yet to be identified in the research literature.

Research on social agency suggests that personalizing a presentation through using "I" and "you" in a script rather than speaking in the third person tends to prime the conversation schema, resulting in greater cognitive effort (e.g., Mayer et al., 2004). This is a technique that students can easily use to create more effective SGM learning products. The notion of social agency lends itself to the idea that using a dialogue structure (such as novice-expert or pro-con) may also prime the conversation schema in a listener.

There exists an impressive body of research (e.g., see Chi, 2000) indicating that deep levels of self-explanation have a very positive effect on learning. For example, Mayer et al.'s (2003) work suggests that if the learner engages in self-explanation, then that will result in greater cognitive engagement. This elevated level of engagement is what Mayer et al. think leads to better learning results. While there may be a number of ways to prime self-explanation, one reasonable design criterion may be to ask students making SGM to contextualize their presentation by connecting the concept to something familiar or relatable to them. This challenge to contextualize will likely activate self-explanation. From the perspective of SGM rather than researcher-generated multimedia, it appears that self-explanation is probably the most important cognitive agent explaining the effectiveness of SGM learning challenges.

Motivational Agents

There are three general recommendations the SGM model suggests: engagement, choice, and relevance. In a recent review of research, Schraw, Flowerday, and Lehman (2001) offered several suggestions on how to increase classroom situational interest, but the strongest predictor of a high-situational-interest classroom can be explained by enhanced levels of student *engagement* (Mitchell, 1993). The primary benefit of using SGM learning challenges is that they tend to encourage students to be actively engaged in their learning. A second suggestion from the research is to offer meaningful choices to students. In some SGM contexts, meaningful choice would seem to be a natural technique to incorporate ("Choose one theory we have studied in class to become an expert in. Your challenge is to teach future students about the essence and relevance of this theory using a multimedia format."). A third suggestion from the research is that relevance is of critical importance (Schraw, Flowerday, & Lehman, 2001). In most SGM learning challenges, students can be asked that their product explain to the target audience the relevance of the specific

concept they are presenting. When students need to explain a concept to others, one of the key factors to creating an effective explanation is to contextualize the information. Thus, for students engaged in SGM learning challenges, the need to contextualize their work probably activates both self-explanation and relevance as priming agents for active learning.

Postproduction

There seem to be two key positive contributions of SGM in the postproduction phase. One of these benefits is mainly for the instructor and future students: Indexing through the creation of a library of multimedia projects provides future students with a library of case studies. Just as filmmakers learn from one another's work, students can learn from the multimedia work of previous students. An easily accessible library of student work can allow future students to assimilate the techniques of past students, thus providing a foundation for developing better and more effective SGMs in the future. The second benefit is usability, wherein the resulting product can be shared and viewed by the current and future students. The ability of multimedia products to be easily shared with other students in the present and future makes an SGM learning challenge more practical and motivating. Just as putting on a theatrical play can enhance motivation because one is putting on a show for a real audience, SGM seems to offer similar benefits through its usability.

EXPLORATIONS INTO SGM

The purpose of this section is to point the reader to key studies on SGM. A relatively large amount of literature has looked at the role of hypermedia and student-generated products, but these studies have been left out because they tend to be text intensive. Instead, studies that were more closely aligned with the spirit of Mayer's (2001) cognitive theory of multimedia are included, that is, studies where the multimedia products represented a better balance and synchronization between the verbal (audio narration) and the visual.

Programming

Probably the best known research that fits under the rubric of SGM was conducted using the LOGO programming language. Papert (1980) wrote extensively about the value of students constructing their own mathematical understanding through this intuitive programming language. More recently, Kafai (1995) and Kafai, Ching, and Marshall (1997) extended the research using LOGO to look at the power of multiple mathematical representations while engaged in software design. As they put it, "While knowledge reformulation is an important feature of learning through software design, personal expression of one's ideas is another" (Kafai et al., p. 118). Their research used fourth and fifth graders who were given the challenge of creating a multimedia learning product that would teach other students about the key concept of fractions. The products students created in this environment displayed a deeper level of understanding of fractions than would normally be expected from the age group. While the research touches many issues, the bottom line is the results indicate that learning to teach via creating SGM learning products seems to result in a deeper understanding of difficult conceptual content.

Spreadsheets

Spreadsheet software has also been used as a multimedia learning tool. Two studies by Mitchell (1997, 2002) looked at the learning-to-teach approach using Microsoft Excel spreadsheet software. The products in both of these studies combined words and images to create learning environments. The words were formatted as text and the images were mainly in the form of interac-

tive raw data that resulted in the end user being able to see dynamic changes in key statistical calculations and graphs. Mitchell (1997) looked at such an approach to understanding foundational statistical concepts, while Mitchell (2002) used a similar approach for students better understanding advanced analysis of variance techniques. Both studies used doctoral students in education as the sample. The results from these two studies supported Benware and Deci's (1984) active-learning hypothesis both in regards to better learning and enhanced motivation through a greater level of engagement. Many of the learning benefits observed in these studies had to do with students developing a deeper connection between statistical techniques and specific research designs, as well as enhancing their interpretive abilities.

Movies

A more generic approach to SGM learning challenges takes the form of movies that integrate visual and audio material. This general approach can more easily be adapted to meet a wider variety of content-learning needs. The most common format for these multimedia movies is as a QuickTime movie (Apple Computer, 2002) since the QuickTime architecture both supports, and can integrate, many different image, video, and audio formats.

In an early study using multimedia, Carver, Lehrer, Connell, and Erickson (1992) identified 16 major thinking skills required of a multimedia designer. They worked with middle-school students and found great support for the development of cognitive skills. One of the conclusions of their study called for the development of "broader theories of assessment that are appropriate for complex tasks like design" (p. 402). Later, Wilson (1999) conducted a study using the middle-school environment in mathematics. Students were put into small groups that created multimedia documents demonstrating their understanding of mathematical problem situations. Wilson found great support for the self-explanation variable: "As one student stated, 'I learned that when you have to teach others you understand better yourself'" (p. 147). Wilson concluded that using multimedia involved a process that cannot be duplicated with more traditional single-media approaches such as writing a paper or creating an illustration. Specifically, Wilson thought there was an important additional level of learning that occurred when students use multiple representations within one product. A recent study by Liu and Hsiao (2002) found that making multimedia was effective in terms of engaging students cognitively and motivationally. They posited that the key to this engagement was putting students in the position of being a designer.

Two recent studies reported on SGM within technology-oriented courses. Steed (2002) looked at senior undergraduates in a course about multimedia and learning. She wanted to know whether peer and target-population evaluations influence the production of student-generated multimedia. Steed's conclusion was that feedback from an audience is a powerful influence on implementing changes to multimedia projects. Neo and Neo (2002) looked at undergraduate students in an interactive multimedia course where they were allowed to design projects of their own choice. Using a survey instrument, Neo and Neo found that students rated the multimedia design process very highly because (a) it was challenging, (b) they were able to be creative thinkers, and (c) they thought the use of multimedia was a more effective approach to presenting their concepts than other alternatives (such as written papers).

Mitchell (2003) looked at SGM learning challenges within two different courses for doctoral students in education (one course called Creativity and the other Motivation). The study looked at how well Mayer's (2001) theory predicted the quality of resulting student products, and whether Benware and Deci's (1984) active-learning hypothesis predicted student motivation. Mayer's theory effectively identified better student products due to the

use of parsimony, audio narrative, personalization, as well as other features of Mayer's model. From a motivational perspective, students appeared to show greater engagement and attention to details, and perceived themselves as co-constructing the curriculum to a level not present in former non-multimedia versions of the same courses.

Two related studies (Andreatta & Mitchell, 2004; Mitchell, Capella, & Andreatta, 2004) took a closer look at the use of SGM with a foundational-level statistics course for doctoral students in education. Students were given two randomly assigned statistical concepts they had to turn into multimedia learning products over the semester-long course. The multimedia products needed to combine audio narration with visual images. Students first completed the audio narrations, received feedback from the instructor, and then went on to complete the full multimedia products. Students found these multimedia learning assessments to be very challenging but more worthwhile than more traditional forms of assessment such as tests or written papers. Especially apparent was the importance of the self-explanation aspect of the SGM model: Since the students were struggling to understand these new statistical concepts, the challenge to clearly and convincingly teach someone else about the same concept pushed them to deepen their own understanding of the topic through the use of analogies, contextualization, worked examples, and other learning devices.

FUTURE TRENDS

SGM learning challenges seem to hold great promise for educators who want to create environments that will both motivate students and push them to learn the target content at a deeper level of understanding. As the practical bottlenecks surrounding the creation of SGM become lessened due to better software and better hardware storage capability, it is likely that more instructors within regular content-driven courses will become interested in this approach. Much of the current research has been in the content area of mathematics. Yet, since creating multimedia is not content specific, it is likely that instructors in more varied academic areas will start using this approach. The educational research community has a large role to play in our understanding of the key learning conditions that need to be in place if SGM learning challenges are to be truly beneficial for student learning. Multimedia is not a magical wand that automatically creates better learners: The support of the research community will be needed so that future instructors can make informed decisions about whether SGM is a worthwhile and practical approach for them to use with their students.

CONCLUSION

Mayer's (2001) cognitive theory of multimedia has been helpful in understanding the basic conditions under which multimedia learning can be effective. His model, which focuses on meaningful outcomes through active learning, can easily be adapted to the special case of SGM. The proposed theory of SGM in this overview suggests that there are both cognitive and motivational factors influencing whether SGM learning challenges will result in students attaining meaningful learning outcomes. An additional potential benefit is the flexibility by which making multimedia can meet the needs of students with a variety of learning styles. Many learners have difficulty learning and communicating in text-dominant environments. Making multimedia allows students to demonstrate their understanding using a much greater variety of communication tools, such as images, that may fit their particular learning style more effectively. While there is a commonly held view today that students learn best by active learning (e.g., see Reigeluth, 1999), the key instructional issue is how to effectively create the conditions that maximize

the probability that each learner will engage in active learning. SGM learning challenges offer one potentially powerful approach to creating the conditions that would prime motivational and cognitive agents, resulting in active learning and, consequently, a deeper understanding of important conceptual content.

REFERENCES

Andreatta, P., & Mitchell, M. (2004). Student attitudes towards statistics instruction using multimedia lectures and assessment. Paper presented at the *Annual Conference of the American Educational Research Association,* San Diego, CA.

Apple Computer. (2002). *QuickTime, version 6* [Computer program]. Palo Alto, CA: Author.

Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal, 21,* 755-765.

Brown, D. (Ed.). (2000). *Interactive learning: Vignettes from America's most wired campuses.* Bolton, MA: Anker Publishing.

Carver, S., Lehrer, R., Connell, T., & Erickson, J. (1992). Learning by hypermedia design: Issues of assessment and implementation. *Educational Psychologist, 27*(3), 385-404.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum.

Chi, M. T. H., Bassock, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Hede, A. (2002). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia, 11*(2), 177-191.

Kafai, Y. (1995). *Minds in play: Computer game design as a context for children's learning.* Hillsdale, NJ: Lawrence Erlbaum.

Kafai, Y., Ching, C., & Marshall, S. (1997). Children as designers of educational multimedia software. *Computers & Education, 29*(2/3), 117-126.

Liu, M., & Hsiao, Y. (2002). Middle school students as multimedia designers: A project-based learning approach. *Journal of Interactive Learning Research, 13*(4), 311-337.

Mann, B., Newhouse, P., Pagram, J., Campbell, A., & Schulz, H. (2002). A comparison of temporal speech and text cueing in educational multimedia. *Journal of Computer Assisted Learning, 18,* 296-308.

Mayer, R. (2001). *Multimedia learning.* Cambridge, England: Cambridge University Press.

Mayer, R., Dow, G., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*(4), 806-813.

Mayer, R., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*(2), 389-395.

Mayer, R., & Massa, L. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833-846.

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*, 427-439.

Mitchell, M. (1997). The use of spreadsheets for constructing statistical understanding. *Journal of Computers in Mathematics and Science Teaching, 16*(2/3), 201-222.

Mitchell, M. (2002). Constructing analysis of variance (ANOVA). *Journal of Computers in Mathematics and Science Teaching, 21*(4), 381-410.

Mitchell, M. (2003). Constructing multimedia: Benefits of student-generated multimedia on learning. *Interactive multimedia electronic journal of computer-enhanced learning, 5*(1). Retrieved from *http://imej.wfu.edu/articles/2003/1/03/index.asp*

Mitchell, M., Capella, E., & Andreatta, P. (2004). Multimedia statistics: A design-based study of the benefits of student generated multimedia for learning in a foundation-level statistics course. Paper presented at the *ED-MEDIA World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Lugano, Switzerland.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92*, 724-733.

Neo, M., & Neo, K. (2002). Building a constructivist learning environment using a multimedia design project: A Malaysia experience. *Journal of Educational Multimedia and Hypermedia, 11*(2), 141-153.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.

Reigeluth, C. (Ed.). (1999). *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. 2). Mahwah, NJ: Lawrence Erlbaum Associates.

Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review, 13*(3), 211-223.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*(1), 23-51.

Steed, M. (2002). The power of peer review in multimedia production. *Journal of Educational Multimedia and Hypermedia, 11*(3), 237-250.

Van Gerven, P., Paas, F., Van Merrienboer, J., Hendriks, M., & Schmidt, H. (2003). The efficiency of multimedia learning into old age. *British Journal of Educational Psychology, 73*, 489-505.

Wilson, M. (1999). Student-generated multimedia presentations: Tools to help build and communicate mathematical understanding. *Journal of Computers in Mathematics and Science Teaching, 18*(2), 145-156.

Zhao, Y., & Frank, K. (2004). Factors affecting technology uses in schools: An ecological perspective. *American Educational Research Journal, 40*(4), 807-840.

KEY TERMS

Active Learning: The three-pronged process of selecting, organizing, and integrating information. The basic idea is that when active learning is optimized, then meaningful outcomes can be more effectively reached.

Cognitive Load: Refers to the limited capacities of the visual and auditory channels in working memory. Since these channels can easily become overloaded, one potential benefit of multimedia learning is being able to present complex information in a format that minimizes potential overloads.

Concise Narrated Animation: This type of multimedia presentation uses synchronized audio narration along with visual material that is concise and uses a meaningful structure (such as a cause-and-effect chain).

Meaningful Outcomes: Operationally measured in most research as transfer or problem solving. More generically meaningful outcomes refer to outcomes that are typically difficult yet essential for students to master.

Multimedia Instruction: A presentation using both words and pictures that is intended to promote learning.

Priming Agent: Refers to any structural feature of a learning environment that increases the likelihood that active learning will take place.

Self-Explanation: Chi et al. (1989) used the term to describe their observation that effective learners established a rational for the solution steps in a problem by pausing to explain the example to themselves. Within multimedia, the working idea is that if a learning experience can prompt the student to be involved in self-explanation, then that will result in greater cognitive engagement.

Social Agency: The idea that social cues can prime the conversation schema. The benefit is that the learner will expend greater cognitive effort when they feel they are part of a conversation than when they are listening to a more formal narration.

Student-Generated Multimedia (SGM): The specific case when students create multimedia products rather than the instructor or outside developer. The purpose of SGM is to serve as a conduit for enhancing student learning of conceptually challenging material.

Student-Generated Multimedia Learning Challenge: The purpose of using an SGM learning challenge is to encourage key priming agents to be activated. The purpose of the structure of such a challenge is to maximize the likelihood that students will achieve meaningful outcomes through the creation of student-generated multimedia products.

This work was previously published in Encyclopedia of Distance Learning, Vol. 4, edited by C. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers, and G.A. Berg, pp. 1693-1702, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.14
An Embedded Collaborative Systems Model for Implementing ICT-Based Multimedia Cartography Teaching and Learning

Shivanand Balram
Simon Fraser University, Canada

Suzana Dragicevic
Simon Fraser University, Canada

ABSTRACT

Information and communication technologies (ICT) have created many new opportunities for teaching, learning and administration. This study elaborates a new embedded collaborative systems (ECS) model to structure and manage the implementation of ICT-based pedagogies in a blended learning environment. Constructivist learning, systems theory, and multimedia concepts are used in the model design and development. The model was applied to a third-year undergraduate multimedia cartography course. The findings show that regardless of student background, implementing effective ICT-based learning pedagogies can be managed using the ECS model.

INTRODUCTION

Integrating information and communication technologies (ICT)—specifically computers, networks, and the Internet—into higher education has created new opportunities for teaching, learning, and administration. Indeed, the role of ICT in the administration of the higher education process has been reflected in national initiatives such as the 1997 Dearing Committee of Inquiry

into Higher Education in the United Kingdom (Dearing, 1997). One of the recommendations of the Dearing Committee was the adoption of national and local ICT strategies to improve the effective and efficient use of resources by U.K. education institutions. Canadian higher education has echoed these strategies and has also increasingly used ICT in the improvement of the quality of distance-education models (Farrell, 1999). The diffusion of information and communication technology into higher education can be attributed to its potential to leverage education processes toward richer and more rewarding learning and management environments (Mitchell, 2002).

In teaching and learning, ICT is a platform on which key learning skills can be efficiently integrated into existing curriculum to boost learner motivation, deepen inquiry, accelerate learning, and widen participation among traditionally isolated groups (Hassell, 2000). Moreover, teaching core ICT skills such as computer operation and programming prepares students to function and succeed in an increasingly information-based society. However, some authors have pointed out that excessive optimism about the micro and mega benefits of ICT in education can develop into broken promises (Selwyn, 2002). These broken promises can adversely influence the adoption of ICT in educational contexts. While most educators agree that ICT has transformed the traditional education process and, hence, demands a new way of thinking, some have pointed out that achieving and verifying useful ICT educational benefits will require strong theoretical evidence, embedded analysis, and research to surmount the associated structural and cultural barriers (Kenway, 1996).

The utility of ICT in providing and retrieving information is of immense value to educators. Instructional designers are now better able to include a range of ICT-based pedagogy into curriculum design and delivery. Many accept that the technology itself does not ensure learning but acknowledge that it enhances traditional instructional systems to deal with modern-day literacy that is a key component of all education goals. Literacy is now generally considered as a multimedia construct (Abbott, 2001). Multimedia improves upon the traditional text and speech formats of interacting with knowledge by integrating other forms of media, such as audio, video, and animations into the learning experience. This has made information more accessible and understandable. But the benefits of multimedia have also come with new challenges. Using multimedia in the classroom is a clear departure from traditional expectations and requires a new mindset and commitment from educators and administrators to ensure effective implementation. Challenges also arise due to the lack of consistent baseline experience to guide the integration of multiple media into the curriculum. Moreover, the wide range of multimedia tools available present a technical challenge to educators who must select instructional technologies to match pedagogical strategies and desired learning outcomes (Abbott, 2001). These challenges demand a flexible and systematic mechanism for managing multimedia tools in traditional learning. Systems theory provides a useful foundation to develop such a management mechanism. In systems theory, the key components of the process are identified and managed separately but as a part of an integrated and functional whole. The resulting systematic structuring ensures that valid models for pedagogy inform the learning process, and that the quality of education is maintained and improved through dynamic interactions between learners and educators.

The utility of ICT in promoting sharing and collaboration among learners is also highly desired. This is reflected in the many content management systems (CMS), such as WebCT (http://www.webct.com), that empower educators to implement synchronous and asynchronous collaborative environments in distance-learning models and in online support for face-to-face instruction or blended-learning models. Socially

mediated constructivist learning theory, where learners explore and discover new knowledge, is the foundation for the collaborative learning paradigm. In face-to-face collaboration, individual and group interactions take place to varying degrees, and finding the appropriate balance is one factor that influence teaching and learning effectiveness (Norman, 2002). Mediating these interactions with technology also presents challenges. Research has shown that non-technology learners in traditional learning settings who do not have access to desired levels of technology support are less willing to use and interact with the learning technology (Watson, Blakeley, & Abbott, 1998). This challenges educators to embed the ICT-based collaborative learning pedagogies into the curriculum structure and design.

The goal of this study is to elaborate on a new embedded collaborative systems (ECS) model for structuring and managing the implementation dynamics of ICT-based pedagogies in a blended learning environment. The specific questions addressed are as follows: How can we engage students in more meaningful learning activities to develop multiple skills of relevance? How can we achieve a useful balance between teacher-centered learning and student-centered learning? The literature on constructivist learning, systems theory, and multimedia education provides the theoretical basis for developing the model. The model was applied to a third-year undergraduate multimedia cartography course of 47 students with no prior knowledge of multimedia and with basic computing skills. The results show that regardless of student background, implementing effective ICT-based learning pedagogies can be managed using the ECS model.

PROMOTING MULTIPLE SKILLS OF RELEVANCE

The focus on the mastery of cognitive and technical skills in the modern-day classroom is a tendency inherited from traditional learning systems. There is now increasing evidence in the workplace to suggest that in the complex problem-solving environment of the real world, the ability to link classroom knowledge with soft skills is a requirement for success. The capability to work in teams, being an enthusiastic and good communicator, infectious creativity, initiative, willingness to learn independently, critical thinking, analytical abilities, self-management, and ethical values are the main soft skills that are highly valued by employers. These new requirements place additional responsibilities on educators to impart knowledge or hard skills together with soft skills in teaching and learning activities so as to prepare learners to function beyond the classroom. This raises the question: How can we engage students in more meaningful learning activities to develop multiple skills of relevance? This question can be examined using a foundation of constructivist learning theory. In this theory, learning is characterized by shared goals and responsibilities, and knowledge is constructed in a discursive environment. Social networking and peer encouragement help motivation and aid individual learning experiences.

Collaborative and cooperative learning have their origins in constructivist learning theory. The goal of collaborative learning is to help learners display individuality and creativity in working with a group toward achieving targets. For collaborative tasks, rewards for achievements are allocated by comparative or normative evaluation systems. In cooperative learning, the focus is on efficiency and effectiveness in achieving a common goal in socially interactive settings (Piaget, 1926; Vygotsky, 1978). In this approach, rewards are allocated based on the quality or quantity of the group product measured against a predefined standard. Although collaborative and cooperative learning share similarities, they differ in their assumptions about competition. Collaborative learning assumes conflict as a part of learning, while cooperative learning tries to minimize

this conflict (Bruffee, 1995). One way to resolve this contradiction is to implement the learning approaches in a way so as to extract the positive learning benefits from each.

BALANCING TEACHER-CENTERED AND STUDENT-CENTERED LEARNING

Implementing multiple learning skills activities requires a balance between teacher-centered and student-centered learning within the contact time limitations of the face-to-face classroom. Thus, efficient course management and structuring become important needs with which to keep track of the evolving course dynamics. Norman (2002) outlined a model that defines the interaction space among a set of agents and objects in the learning process. In the model, two sets of agents (instructors and students) and two sets of objects (course materials and course products) overlap to form a complex interaction space. This results in six intersecting areas that form regions where a combination of two or more agents or objects exists. The usefulness of this interaction model is that it shows the variety of interacting elements that require management during the learning process. But while the model provides a comprehensive description of the interactions, it does not deal explicitly with how to balance these dynamic interactions during the learning process. This raises the question: How can we achieve a useful balance between teacher-centered learning and student-centered learning? This question can be examined using a systems theory approach.

Systems theory can be used to manage the instructional tools used to facilitate teaching and learning among the agents. In this way, the theory guides the efforts in balancing the load between student-centered and teacher-centered learning. The theory considers the teaching and learning process to be composed of a set of tightly interrelated pedagogies that can be used to communicate and deliver educational content (Bertalanffy, 1969). Based on the systems approach, together with the constructivist paradigm, a wide range of possible pedagogies can be identified. Examples of these pedagogies include learning contracts, brainstorming, debate, observation, simulation, case study, discussion, and forum. By integrating these approaches systematically, an equitable balance between teacher-centered learning that communicates knowledge and student-centered learning that integrates all levels of Blooms Taxonomy (knowledge, comprehension, application, analysis, synthesis, and evaluation) can be achieved.

MULTIMEDIA CARTOGRAPHY TEACHING AND LEARNING

The use of computer-based technologies in geography teaching and learning has a long and rich tradition (Gold et al., 1991). This stems from the influence of the quantitative revolution on many areas of the subject. Spatial information studies (encompassing geographic information systems and science, remote sensing, digital cartography, and spatial analysis) are a product of that quantitative influence. Over the last decade, ICT and specialized research software for geography, in general, and spatial information studies, in particular, has caused many changes to the community of research, learning, and teaching practices in these areas. Spatial information studies educators are now battling with how best to balance knowledge transmission with the necessary software practice in the learning process. The "cookbook" approach of traditional lectures and independent student learning of computer skills are two extremes in the learning spectrum of an increasingly computer-driven curriculum. Clearly, any solution must deal with establishing structures for an equitable distribution of the

pedagogies across the curriculum and focus the pedagogies on skills students need for success in further studies and the workplace.

Cartography encompasses the art, science, and technology of making maps and requires diverse technical and creative skills for effective practice. The use of multimedia in cartography education serves two interrelated functions: as an instructional tool and as a product development tool. Instructional frameworks to incorporate multimedia-based instruction into the curriculum must be consistent with existing theories of teaching and learning. This has been emphasized by a number of researchers (Alessi & Trollip, 2001; Benyon, Stone, & Woodroffe, 1997; Ellis, 2001; Najjar, 1996). The multiple representation (MR) framework allows the inclusion of knowledge domains within multimedia (Kinshuk & Patel, 2003). The MR approach involves the selection of multimedia objects, navigational objects, and the integration of multimedia objects in the representation of the knowledge domains. Teaching strategies and styles are also important factors in multimedia learning, as they impact learning. The benefits of multimedia education include improved learning retention, portability, modularity, enhanced visualizations, efficiency in instructional design, and learning consistency (Hede, 2002; Yildirim, Ozden, & Aksu, 2001).

The use of multimedia authoring tools in designing course products enables learners to develop and construct enhanced mapping products. This forms the basis of multimedia cartography, in which the paper map is transformed into an enhanced digital map that integrates multiple media to communicate visual and oral expressions of spatial information to the map reader (Cartwright, Peterson, & Gartner, 1999). These multimedia maps are accessed through CD-ROM, the Internet, or specially designed Web-mapping services. The benefits of multimedia maps include dynamic and multifaceted representation of space and time, superior map production and dissemination, improved information and knowledge transfer, and greater map accessibility.

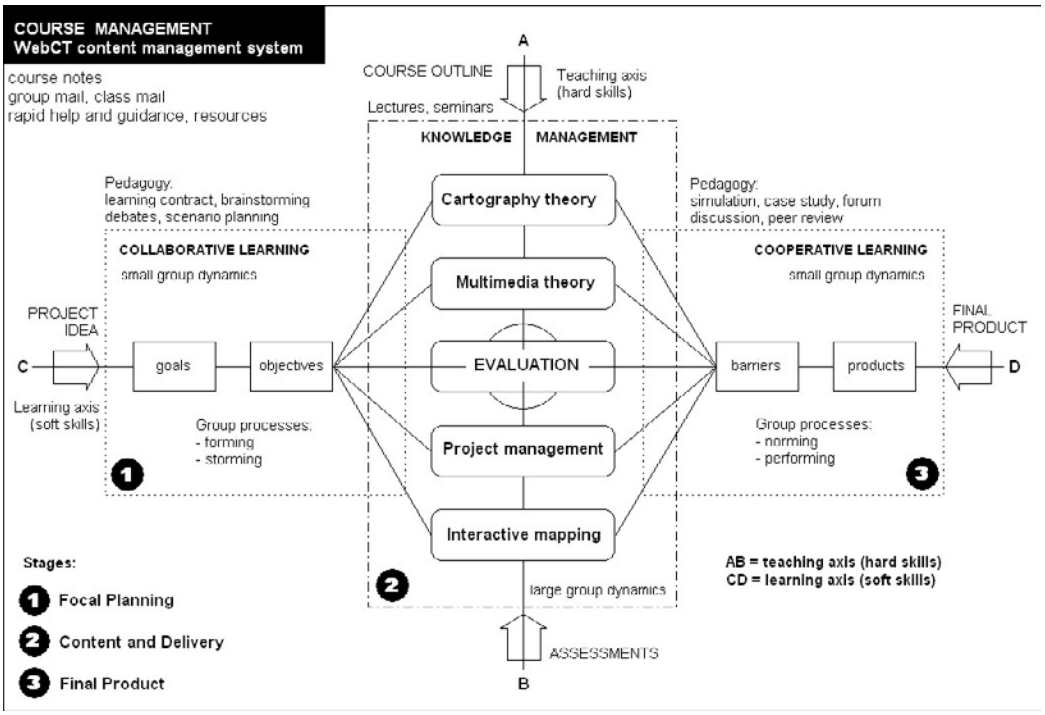
EMBEDDED COLLABORATIVE SYSTEMS MODEL

The embedded collaborative systems (ECS) model is designed based on principles from constructivist learning theory and systems theory. The goal of the model is to improve the quality of student learning in the face-to-face classroom. This is achieved with a focus on the development of multiple learning skills and on independent learning. The model's structure is shown in Figure 1.

The collaborative, knowledge management, and cooperative working spaces are three distinct overlapping interaction spaces that are defined in the model. The overlapping structure strengthens the process and provides connectivity among the stages of focal planning of projects, delivery of course content, and preparation of final course products. The knowledge management workspace occurs in the classroom, where all students receive the same content through lectures and seminars. The collaborative and cooperative workspaces occur during small group laboratory sessions or alternatively in informal meetings among students.

The teaching (AB) and learning (CD) axes serve as both workflow and information flow pathways in the model. These axes control the levels of hard and soft skills that are integrated in the learning experience. The hard skills or teaching axis deals with the substantive course content. This content is normally stipulated by institutional curriculum policies and is implemented using traditional pedagogical tools such as lectures, seminars, and panels. The knowledge management phase of the process is implemented in large groups to encourage critical thinking and discussions. Students develop individual and active learning habits during all stages of the hard

Figure 1: The embedded collaborative systems (ECS) model



skills implementation. The course outline and content together with the assessment requirements drive the nature of the interactions that occur along the teaching axis.

The soft skills or learning axis characterizes the collective and social interaction experiences of students working to achieve targets in a group environment. Examples of pedagogies that can be used involve group projects, learning contracts, brainstorming, simulation, forum, discussions, and case studies embedded in real problem-solving contexts. The intersection of the learning and teaching axes provides an opportunity for formative evaluation. Formative evaluation is an important component of the process, as with it, we are able to establish how students are integrated into the learning experience and how satisfied they are with the learning environment. Formative evaluations include interviews and survey questionnaires, and corrective action is immediately implemented to control any identified imbalances. Evaluations take the traditional form of cognitive assessments using normative testing instruments.

The use of multiple pedagogies provides students with the opportunity to experience deeper learning as they master new concepts by manipulating and refining previous knowledge. The pedagogical tools and the instructional medium appropriate for each stage of the learning process are described and explained in Table 1. The flexibility of ECS model allows students to pursue topics of general interest during the group projects. This supports the assumption that learning is a lifelong process, and learners have a role in designing what they learn.

Achieving a balance between student-centered and teacher-centered learning is inherent in the ECS model. During the initial stages of the model implementation, teacher-centered cognitive learning is at a high level, whereas student-centered learning is at a low level (Figure

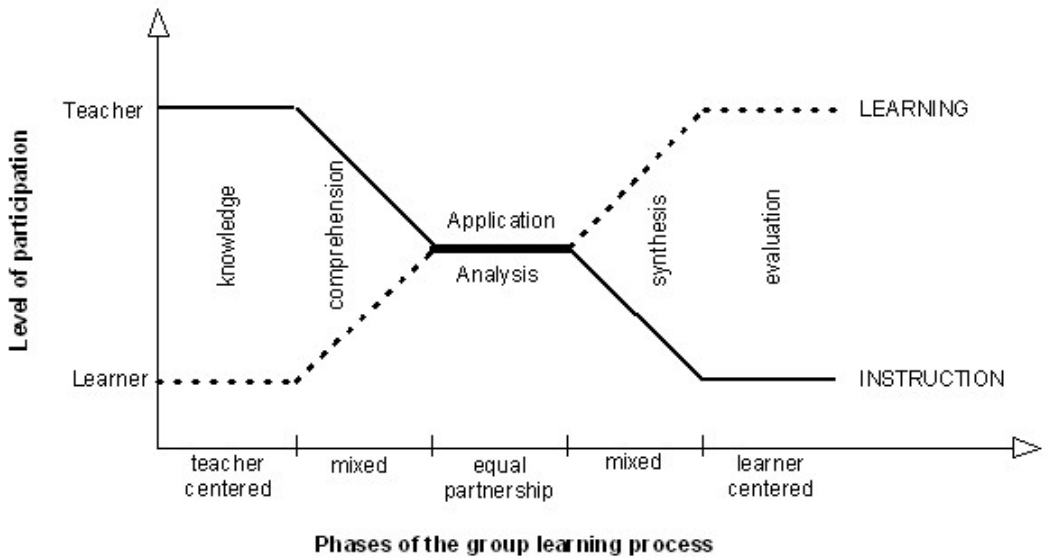
Table 1: Pedagogies and instructional media used in the ECS model

Pedagogy	Description of the Pedagogy	Instructional Media	Stage in the ECS Model	Targeted Skills
Presentation	Instructor-centered lecture notes and student-centered communication of project results	Multimedia, videotape, graphic visuals, text	1,2,3	Hard and soft skills
Discussion	Exchange of ideas and opinions among students–students and teacher–students; discussions are guided by reflective questions	Graphic visuals, text	1,2,3	Soft skills
Demonstration	Instructor-centered presentation of example of skills to be learned; use of expert to present case study	Software, the Internet, graphic visuals	2	Hard skills
Drill and Practice	Exercises such as assignments to reinforce skills	Software, text	2	Hard skills
Tutorial	Individual learning through practice and feedback	Software, text, the Internet	1,2,3	Hard and soft skills
Group work	Small group work on defining projects and allocating tasks	Software, graphic visuals, text, multimedia	1,3	Soft skills
Simulation	Experimentation with small version of reality that is to be described and understood	Software, graphic visuals, text, multimedia	1,3	Soft skills
Gaming	A user-friendly environment for testing specific rules and their effects on determined goals	Software, text	3	Soft skills
Discovery	Problem solving through trial and error	Software, graphic visuals, text, multimedia	1,3	Soft skills
Problem solving	Applying skills to find solutions to real problems	Software, videotape, graphic visuals, audio, text, multimedia	3	Hard and soft skills

2). At the beginning of the collaborative stages, a progression is seen through a fuzzy period of mixed learning toward an equal partnership in learning between teacher and learner. Thereafter, students gradually become equipped with the skills and motivation to undertake active and independent learning. Eventually, instruction is replaced with independent learning, as the full spectrum of Blooms Taxonomy is covered. The timing of the introduction of the collaborative and cooperative tasks coincides with the stages of the learning processes shown in Figure 2.

The ECS model is optimized for blended learning environments, where face-to-face instruction is supported and complemented by online instruction. Content management systems (CMS) such as WebCT (http://www.webct.com) offer comprehensive administration tools with which to deploy complex pedagogies that can emerge using the ECS model. The use of systems theory allows the educator to identify the major pedagogical components that will best achieve the desired learning outcomes. In addition, systems theory integrates the knowledge and pedagogy of the process through rigorous alternatives assessments. Each separate component of the process is analyzed for relevance and then integrated to consolidate and expand individual learning. This

Figure 2: The dynamic phases of the group learning process



framework structures the learning environment, provides a mechanism for understanding inter-relationships, and provides task balancing and process management benefits among others. The central aspect is that a systematic framework for group interactions is established that allows teams to define roles, define protocols for independent working, and devise strategies for individual accountability.

APPLICATION OF THE ECS MODEL

Cartography Course Background

The multimedia cartography course used to test the model was at the third-year undergraduate level and consisted of 47 students. The total contact duration was 13 weeks. Two hours of formal lectures and two hours of computer lab work were compulsory, guided sessions each week. The lectures were delivered to all students at the same time, while the computer labs were conducted in three sessions with not more than 20 students attending per session. The rationale for multiple lab sessions was to ensure that students had access to computer resources and were able to receive individualized attention from the teaching assistant. The classroom and lab settings exposed students to both teacher-centered instructions and learner-centered instructions. Students initially had little knowledge of multimedia concepts, cartography theory, or relevant software tools. But this situation was ideal for investigating the ECS model for learning effectiveness among students and the management of the learning process.

Designing learning structures that stimulate and promote enhanced student motivation is perhaps the most crucial aspect of learning (Edstrom, 2002). Identifying motivations allows instructors to develop strategies for redirecting student goals toward more meaningful and rewarding learning experiences. A questionnaire survey was implemented at the beginning of the cartography course to determine student motivation and rationale. The open-ended anonymous question:

"What do you expect to achieve by attending this course?" provided valuable responses (Table 2). Learning about Internet mapping and cartography principles was the most frequent statement given by students who responded. This indicated that student motivation was generally aligned with the course objectives, and hence, more time would be available for the instructor to focus on preparing engaging content. As is expected, some students were interested in software learning to improve their job prospects and others on obtaining the necessary credits toward graduation.

The open-ended anonymous question: *"What can the instructor and teaching assistant do during the lectures and labs to make you learn better?"* indicated that the most frequent expectation was for clearly explained example-based content (Table 3). The information obtained from the two questions guided the selection of pedagogical components in the ECS model, so that the learning process was balanced by student expectations and institutional curriculum policies.

Content Structuring and Knowledge Management

The first 4 weeks were dedicated to formal lectures and guided practice on the use of software tools. Moreover, cases were analyzed and best practices extracted such that students became familiar with general practices in the subject area. This forms the knowledge management phase of the model, in which learning proceeds through incremental steps, and individual learning is emphasized. The subsequent weeks were structured so that knowledge management at an individual level and collaborative project at a group level reinforced each other for an enhanced learning

Table 2: Motivation of students in the multimedia cartography course

What do you expect to achieve by attending this course?	**Frequency of Statements (%)** (Number of Statements = 47)
Better understanding of mapping on the Internet	13 (27.7)
Expand my knowledge of cartographic techniques	12 (25.5)
Greater familiarity with the software to be used	7 (14.9)
Academic credits and knowledge	5 (10.6)
Others	10 (21.3)

Table 3: Students' suggestions for a better learning environment

What can the instructor and teaching assistant do during the lectures and labs to make you learn better?	**Frequency of Statements (%)** (Number of Statements = 46)
Give clear and concise explanations	8 (17.4)
Provide many examples during teaching	7 (15.2)
Present the materials at a reasonable pace	4 (8.7)
Make the content relevant and interesting	3 (6.5)
Give well-organized lecture notes	3 (6.5)
Make notes available ahead of lectures	3 (6.5)
Others	18 (39.1)

experience. Formal lectures included concepts related to cartography, multimedia, Web mapping and project management theory (Figure 1). The focus of the lectures was on case studies, and students were exposed to analytical, application, creative, communication, social, and self-analysis skills (Easton, 1982). Moreover, students were able to discuss their views freely and to listen to the views of peers. The group work and peer support operated both as additional instructions for students and as a forum for wider discussions within the course framework.

Of significance in this stage is the concept of Web-based mapping, which involves some level of computer networking knowledge (Figure 3). In a Web-mapping multimedia application, a digital map, once created, becomes a dynamic index to multimedia content. The map is hosted on a Web server, and a map server provides a dynamic link to a database to allow end users to query and interact with the map in the browser window. Although the learning curve for this particular type of mapping technology is steep, it was surprising to find that students were extremely motivated and committed to learning the software. Informal interviews revealed that the general source of this motivation came from the structuring of the learning outcomes at each stage of the process and the out-of-class support and help provided by the teaching assistant. Students were more committed and motivated when they could control how and when they learned.

Accessing notes and supplementary materials before lectures ensures that students concentrate on synthesis and analysis rather than on note taking. The new electronic media make it easy to provide additional readings based on student needs, and the online environment provides a social space for continuous conversations and support among peers. Optimal learning occurs when students share knowledge among peers in a community of practice where ideas are evaluated and adapted. In order to manage the implementation of the model, the content management tool WebCT was used for managing mailing lists, discussions, and presentation of knowledge content (Figure 4). The real-time facilities of the management tools were useful in fostering the "community spirit" outside of the classroom setting. The final multimedia cartography atlas products developed and

Figure 3: Levels of use of relevant software

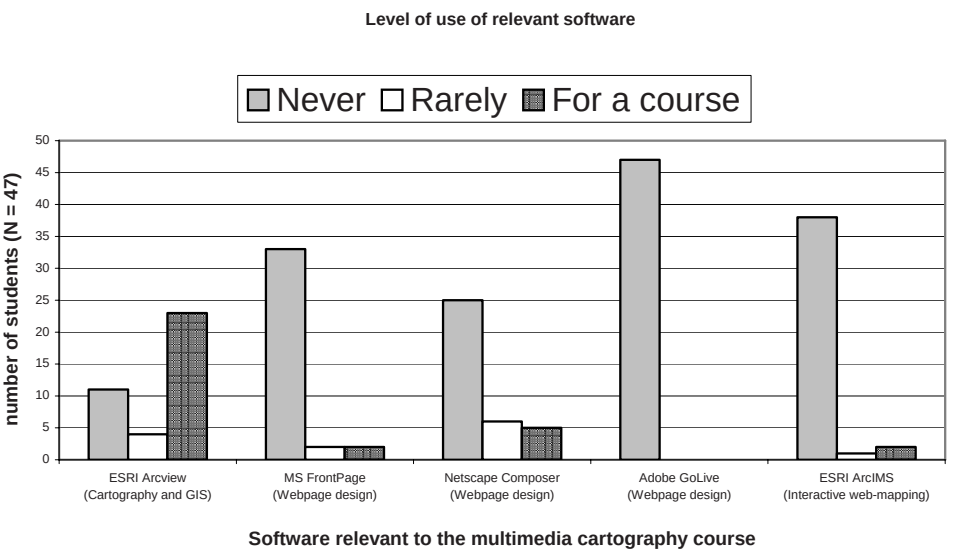
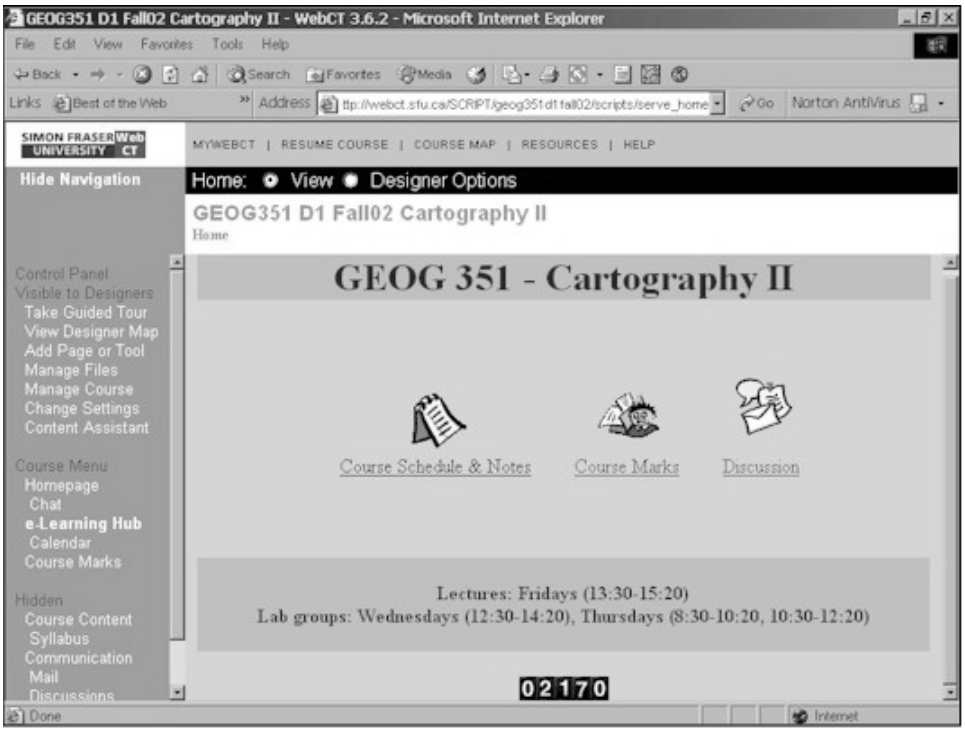


Figure 4: Model implementation using the WebCT content management system



implemented by students are documented on the Web (http://www.sfu.ca/geog351fall02/).

Collaborative and Cooperative Learning

In the collaborative learning exercise, learners were divided into nine groups (average of five students per group, some randomly and some based on individual preference), and step-by-step guidelines were given to each group on the final product to be achieved (designing and creating a professional multimedia atlas), and resources (books, journal papers, Internet Web sites) available for unraveling what is to be achieved. Tasks included mastering sets of specific learning objectives and finding ways to transfer that learning to the class as a whole. There was no briefing to students on group roles. Students reported that the first set of group discussions was difficult. This was expected, but students needed to learn how to cope in a new and unfamiliar social and learning environment. The leadership role was usually assigned to the student with much to say. That leader then assigns tasks, facilitates the discussions, and ensures that meaningful results emerge from the discussions. The weekly meetings provided time for critical reflections and perusal of new materials, ideas, perspectives, and further research. Each week, the instructor met with the group to evaluate problems and progress and to offer solutions strategies. Ideas were not imposed on the groups, and this was much appreciated by the learners. Moreover, learners agreed in principle to abide by "moral and ethical" conduct during the course.

Students were involved in the initiation and definition of a relevant project. Support was provided in the form of guidance about the format of the final products to be produced, potential areas for projects, course aims and expectations, time schedules, data resources, and project

management information. In the design stage, students were encouraged to develop a concept sketch of their product and to begin the process of tool selection and task allocation. There were many opportunities for students to consult with the course instructor and teaching assistant during this stage and, indeed, for the entire course. All members of the group were encouraged to participate in the creation, assembly, and layout of content so as to ensure a uniform individual learning experience. Internet resources and books were provided to help further the process. Each group identified a member to coordinate their activities and to maintain close liaison with the instructors. The final course products were presented to peers for review, and comments were gathered by the instructor and given to each group. This feedback was useful in improving the quality of the products and establishing a standardized level. This peer review also introduced critical and reflective practices into the process (Bazeli & Robinson, 2001).

The use of information technology together with communicating and working with peers was identified by students as contributing to the success of the project. There were instances in which some students were more focused on the technological tools and less on the content. The group-learning format ensured that group members provide focus and guidance to individuals. Evidence of this peer learning was reflected in the sophistication of the cartographic products and how these products quickly diffused and were adopted by other groups. The professionalism of the final products helped in motivating students toward greater learning and explorations.

Assessment and Evaluation

Evaluations have substantial gains for individual projects and progress in the field. Moreover, the use of new ICT technology has resulted in curriculum changes and may require new ways of evaluating students. In the collaborative modes of learning, the focus is on teamwork and communication skills, and appropriate measures of these are needed. An interim way of dealing with this is to use outside evaluators and student questionnaires. Also, multiple levels of evaluation can yield richer feedback from external, internal, and peer sources. Logging usage statistics and student interviews is another way to identify features of the course that are good and what improvements can be made so that future refinements of the process can be made.

Assessment materials were returned to students quickly, showing where improvements can be made. The exercise involved students preparing work individually and bringing it to the group. The small groups then integrated the materials using e-mail and face-to-face meetings. The small group work was then shown to the groups in a large-format presentation. The assessment also allows educators to learn about their work in a critical and reflective way so that rapid improvements can be instituted for the benefit of learners (Gerber, 2002). The assessment for the cartography course was comprised of individual assignments, group mini-project presentations, individual participation, examination, and production of a final working group electronic atlas.

All students completed a questionnaire during the formal group presentations, and some students were randomly interviewed at the midway point in the course to obtain feedback toward formative evaluations. In the group presentations, each student was required to judge the presentations of the others on a 5-point verbal scale ranging from poor to excellent. Moreover, reasons for the judgments were also requested. The general trend of the responses was toward the favourable end of the judgment scale and was justified by the respondents on two main grounds—the sophistication of the tools and techniques used for creating the atlases and the non-duplication of these techniques across the groups. Students clearly indicated that these were attributed to the efficient small-group work and the collaborative

settings in which they occurred. However, one shortcoming identified was the lack of time. While this was unavoidable given the constraints of the semester and curriculum, techniques and tools for time and project management were again reinforced such that facilities for handling this shortcoming were available to them beyond the course. Another shortcoming identified was the variation in skills within the groups. Although students recognized the difficult logistical problems that this can cause, they nevertheless felt that the group's experience would probably have been more rewarding with balanced skills. One way to deal with this is to make greater use of learning styles and skills inventory to categorize students into the small project groups. However, this will demand a trade-off between efficiency in the course logistics and effectiveness in implementing the ECS model. The two comments below characterize the general attitude of students:

"The projects and presentations overall were very impressive and obviously well thought out. The presentations give an overall view of the work-effort placed within each group."

"In general I would like to say that all the atlas were very different concerning layout and information, but most of them were really very good."

A summative evaluation in the later stages of the course elicited responses on the following questions using a 4-point scale (4 being most favourable):

- The assignments and lectures were [unrelated—well related]; mean score = 2.61 (n = 36)
- The exams and assignments were on the whole [unfair—fair]; mean score = 2.81 (n = 36)
- The marking scheme was on the whole [unfair—fair]; mean score = 2.92 (n = 36)

These results from the summative evaluations are inconclusive. While they indicate a general positive weight to the statements, the aggregations of different student backgrounds and experiences makes any interpretation uncertain. The issue of a learner capability to judge curriculum content and implementation is still unresolved in the literature. Nevertheless, the informal interviews and the level of accomplishment in the final atlas products provide strong indication that the collaborative learning process, as implemented using the ECS model, was indeed effective in managing and task balancing the components toward the intended products.

FUTURE TRENDS

The further development and integration of ICT into multimedia cartography education is dependent on three factors: access to ICT tools, instructors' knowledge of effective ICT use, and more studies on the benefits of ICT and multimedia in student learning. Access and instructor knowledge are issues best handled from the wider policies and practices of higher education administration. Systematic research is needed to further establish the role of ICT in learning.

The software and hardware needs for geography education are enormous. Centralized servers for demonstrating and hosting Web-mapping services, the multiplicity and constantly changing software tools, and the need to redesign current computer laboratories to accommodate collaborative group learning are some of the central considerations that will influence the wider adoption and diffusion of an ICT in the geography curriculum. A troubling issue for multimedia cartography teaching and learning is software licensing arrangements that can sometimes be a barrier to using certain software tools in the learning process. This, in some ways, dictates the eventual skills that students can achieve. Technology providers will need to seriously

consider pricing mechanisms so that academic institutions are better able to afford and maintain basic technological infrastructures to implement core teaching and education programs. There has been some progress in this area, with mechanisms such as university campus licensing that enable widespread use of some software tools for teaching and research.

With the gradual expansion of the home as a center of learning, arrangements for students to use university resources at home promise to be a major issue, especially with respect to copyrights and off-campus licensing agreements. University libraries hold a key position in this regard. Already, electronic books, or e-books, are a common feature of many western university library catalogs, and there has been growing evidence to suggest that some of the more progressive university libraries have already begun to redefine their roles as information gateways to act as the intermediary between the user and information (Dowler, 1997). Electronic data archives, multimedia reuseable learning object databases, subject portals, and continuing skills training for students are ways libraries have begun to accept their changing roles in university teaching and learning. A common thread in all the transformations has been the impact that ICT has brought to the university and classroom with respect to administration, teaching, and learning.

Existing models of multimedia cartography teaching and learning have been mostly descriptive. These models have been useful in understanding the mechanisms in operation and in enabling comparisons to be made across different learning contexts. The results from these studies have enabled educators to generally conclude that ICT and multimedia have positive benefits for learning. However, not much is known about the critical factors and how they influence learning. Systematic investigations of predictive models in diverse contexts provide the next steps for understanding the factors of ICT and multimedia that affect learning. Following along this line will be new learning tools, with which intelligent agents will guide learners through knowledge nodes and learning activities using hypermedia and multimedia in much the same way as the intelligent help assistant acts in the Microsoft Office software products.

CONCLUSION

The ECS model presented is based on a holistic perspective of learning as complex interactions between multiple agents, physical and social spaces, and instructional technologies. Although the model can be used in hypothesis testing, the main goal is to provide instructional designers and educators with a tool for managing the main factors that need to be considered when designing ICT-based pedagogies. This model provides the framework for good instructional and course-structuring design that takes into account the diversity of learner styles and provides engaging interactions among students.

The use of ICT and multimedia pedagogy in cartographic education is still in the early stages of understanding and development. There are numerous possibilities and pitfalls. But given the early stages of diffusion of multimedia tools in education, the current focus among practitioners is on developing strategies and standardized protocols to produce effective multimedia components that blend engagement and entertainment into a single learning environment. Moreover, collaborative processes aid the pedagogical move toward student-centered learning.

The embedded collaborative systems model was developed to structure and understand the dynamics involved in the implementation of multiple learning skills activities. The implementation involved 47 students in a multimedia cartography course. The course was conducted in a blended learning environment, and discursive group learning was the foundation of the learning experience (Thorne, 2003). Each group defined

project content, prepared a proposal, defended their proposal in front of their peers in a formal conference-type presentation, received feedback from peers, and used the feedback to improve their group's project. Also, the other groups judged each group on presentations. This forms the cooperative phase, where individuality and group opinions are merged for consensual learning.

In summary, the issues in this study, namely, how to implement effective (content and experience) multimedia cartography training and education to learners of diverse backgrounds, was addressed by the development and testing of a systems model for integrating the multiple facets involved in the education and training process. Within the systems model, the collaborative and cooperative learning strategies were integrated to promote individual and group development for effective multimedia cartography education and product development. The benefits of the ECS model include the following:

- Improves connectivity among the actors by embedded and continuous interaction
- Cultivates an attitude of independent learning through peer guidance and motivation
- Integrates multimedia information, thereby catering to a range of learning styles
- Provides ownership of the learning process through group and individual project management
- Develops individual social and learning skills and accountability

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the following sources: an International Council for Canadian Studies (ICCS-CIES) Scholarship and a Department of Geography (Simon Fraser University) Teaching Assistantship to S. Balram; and a Simon Fraser University President Research Grant to S. Dragicevic. The comments of Dr. David Kaufman, LIDC, Simon Fraser University are gratefully appreciated. The authors thank two anonymous referees for their comments and suggestions toward improving an earlier draft of the manuscript.

REFERENCES

Abbott, C. (2001). *ICT: Changing education*. London; New York: Routledge Falmer.

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development*. Boston, MA: Allyn and Bacon.

Bazeli, M. J., & Robinson, R. S. (2001). Critical viewing to promote critical thinking. In R. Muffoletto (Ed.), *Education and technology: Critical and reflective practices* (pp. 69–91). Cresskill, NJ: Hampton Press.

Benyon, D., Stone, D., & Woodroffe, M. (1997). Experience with developing multimedia courseware for the World Wide Web: The need for better tools and clear pedagogy. *International Journal of Human–Computer Studies, 47*, 197–218.

Bertalanffy, L. v. (1969). *General systems theory; Foundations, development, applications*. New York: G. Braziller.

Bruffee, K. A. (1995). Sharing our toys: Cooperative learning versus collaborative learning. *Change*, (January/February), 12–18.

Cartwright, W., Peterson, M. P., & Gartner, G. F. (Eds.). (1999). *Multimedia cartography*. Berlin; New York: Springer.

Dearing, R. (1997). *The National Committee of Inquiry into Higher Education*. Retrieved April 10, 2003 from the World Wide Web: http://www.leeds.ac.uk/educol/ncihe/

Dowler, L. (Ed.). (1997). *Gateways to knowledge: The role of academic libraries in teaching, learning, and research*. Cambridge, MA: MIT Press.

Easton, G. (1982). *Learning from case studies*. Englewood Cliffs, NJ: Prentice Hall International.

Edstrom, K. (2002). Design for motivation. In S. Hailes (Ed.), *The digital university: Building a learning community* (pp. 193–202). London; New York: Springer.

Ellis, T. J. (2001). Multimedia enhanced educational products as a tool to promote critical thinking in adult students. *Journal of Educational Multimedia and Hypermedia, 10*(2), 107–123.

Farrell, G. (1999). The development of virtual institutions in Canada. In G. Farrell (Ed.), *The development of virtual education: A global perspective* (pp. 13–22). Vancouver, Canada: The Commonwealth of Learning.

Gerber, R. (2002). Understanding how geographical educators learn in their work: An important basis for their professional development. In M. Smith (Ed.), *Teaching geography in secondary schools: A reader* (pp. 293–305). London: Routledge Falmer.

Gold, J. R., Jenkins, A., Lee, R., Monk, J., Riley, J., Shepherd, I., & Unwin, D. (1991). *Teaching geography in higher education: A manual of good practice.* Oxford, UK; Cambridge, MA: Basil Blackwell.

Hassell, D. (2000). Issues in ICT and geography. In T. Binns (Ed.), *Issues in geography teaching* (pp. 80–92). London; New York: Routledge.

Hede, A. (2002). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia, 11*(2), 177–191.

Kenway, J. (1996). The information superhighway and post-modernity: The social promise and the social price. *Comparative Education, 32*(2), 217–231.

Kinshuk, & Patel, A. (2003). Optimizing domain representation with multimedia objects. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and practice* (pp. 55–68). London and Sterling, VA: Kogan Page Limited.

Mitchell, B. R. (2002). The relevance and impact of collaborative working for management in a digital university. In S. Hailes (Ed.), *The digital university: Building a learning community* (pp. 229–246). London; New York: Springer.

Najjar, L. J. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia, 5*(2), 129–150.

Norman, K. (2002). Collaborative interactions in support of learning: Models, metaphors and management. In S. Hailes (Ed.), *The digital university: Building a learning community* (pp. 41–56). London; New York: Springer.

Piaget, J. (1926). *The language and thought of a child.* London: Routledge & Kegan Paul.

Selwyn, N. (2002). *Telling tales on technology: Qualitative studies of technology and education.* Aldershot, Hants, England; Burlington, VT: Ashgate.

Thorne, K. (2003). *Blended learning: How to integrate online & traditional learning.* London; Sterling, VA: Kogan Page.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Watson, D., Blakeley, B., & Abbott, C. (1998). Researching the use of communication technologies in teacher education. *Computers and Education, 30*(1–2), 15–21.

Yildirim, Z., Ozden, M. Y., & Aksu, M. (2001). Comparison of hypermedia learning and traditional instruction on knowledge acquisition and retention. *The Journal of Educational Research, 94*(4), 207–214.

This work was previously published in Interactive Multimedia in Education and Training , edited by S. Mishra AND R.C. Sharma, pp. 1-28, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.15
Multimedia Evaluations Based on Cognitive Science Findings

Eshaa M. Alkhalifa
University of Bahrain, Bahrain

INTRODUCTION

Multi-media systems waltzed into the lives of students and educators without allowing for the time required for the development of suitable evaluation techniques. Although everyone in the field is aware that judging this type of teaching software can only come through evaluations, the work done in this regard is scarce and ill organized. Unfortunately, in many of the cases the evaluation forms were just filled in by instructors who pretended they were students when they went through the tutorial systems (Reiser et al., 1994).

BACKGROUND

In the early days, some researchers regarded the evaluation of the program's functional abilities and efficiency to be important, so they defined them as formative evaluation, and they also defined the effectiveness of the system as summative evaluation (Bloom et al., 1971; Scriven, 1967).

Others believe the evaluation of the system is unimportant, so they focused on the latter by comparing student performance in pre- and post-test questionnaires prior to and following the use of the system, learning style questionnaires that targeted their learning preferences and a subjective questionnaire that investigated whether students like the system (Kinshuk et al., 2000). Unfortunately, many of the pre- and post-tests resulted in no significant1 differences in student grades when multi-media is compared to classroom lectures or to carefully organized, well-illustrated textbooks (Pane et al., 1996). These disappointing results caused researchers to question whether or not the correct evaluation questions are being asked; for example should the test be of interactivity versus lack of interactivity, or should one compare animation with textual media (McKenna, 1995)? If Pane et al. (1996) were aware of the work done by Freyd (1987), who studied the cognitive effects of exposing subjects to a series of still images to find that they are equivalent in the reactions they elicit to being exposed to a moving picture,

then perhaps they would not have asked whether animation is equivalent to a textbook with carefully set images of all stages.

Since the problem that arose is the evaluation question, researchers continued to alter it in order to recognize what should be emphasized. Tam et al. (1997) proposed a three-part evaluation procedure that includes peer review, student evaluation as well as pre- and post-testing (Tam et. al., 1997). They were not able to get rid of the pre- and post-test evaluation, as it is the primary test for how much learning was achieved, and they still got no significant differences.

At this stage, researchers recognized that evaluations did not target the appropriate level of detail, so Song et al. (2000, 2001) presented empirical support that animation helps reduce the cognitive load on the learner. They also showed that multi-media is more effective in teaching processes than in teaching conceptual definitions, while textual presentations are better at the latter. However, all this was done in very limited test domains that lacked the realistic world of an educational system. Albalooshi and Alkhalifa (2002) implemented some of these ideas in addition to offering both textual representations and animations within the same screen to students. This supports individual learning preferences while offering multi-media systems as a cognitive tool. Such a tool requires an evaluation framework that is well informed of the justification behind its design and the way its main modules interact.

A 3-DIMENTIONAL FRAMEWORK FOR EVALUATION

In the reported cases, most of the evaluated systems failed to reflect their true abilities because some aspects of the design or effects were neglected. Consequently, a complete framework of evaluation is required to take into account all issues concerning the software and the learning process. Evaluation questions can be channeled into three main dimensions of evaluation that could then be subdivided into the various methods that form possible criteria that guide the evaluation process.

The framework will be explained through a case study that was performed of a data structure tutorial system (DAST) that was developed and evaluated at the University of Bahrain (AlBalooshi & Alkhalifa, 2003). The process started with a pre-evaluation stage, where students were all given a test and then were divided into groups of equivalent mean grades. This was done to allow each group to have members of all learning levels.

Then the pre- and post-tests were written to ensure that one set of questions mapped onto the next by altering their order while ensuring they include declarative questions that require verbalization of how students understand concepts as well as procedural questions that test if students understand how the concepts can be applied. Last but not least, a questionnaire was prepared to allow students to highlight what they regard as any weak areas or strong areas based upon their interaction with the system. The evaluation procedure for students is shown in Figure 1. Educators were also asked to fill in an evaluation form as experts.

ANALYSIS OF RESULTS

First of all, student grades were analyzed using the Analysis of Variance (ANOVA) test. This test allows the evaluation of the difference between the means by placing all the data into one number, which is F, and returning as a result one *p* for the null hypothesis. It will also compare the variability that is observed between conditions to the variability observed within each condition.

The statistic F is obtained as a ratio of two estimates of students' variances. If the ratio is sufficiently larger than 1, then the observed differences among the obtained means are described as being statistically significant. The term "null

Table 1. A three-dimensional framework of evaluation

1st Dimension: System Architecture

This dimension is concerned with the system's main modules, their programming complexity as well as their interactions. Evaluation within this dimension should be performed in any or all of the following methods:

- Full description of system modules and complete check of interaction.
- Expert survey of the system filled by experts or educators.
- Student evaluations to consider their perspective of the system.
- Architectural design must be based on cognitive science findings rather than chance.
- Everything else concerning the system design such as cost analysis and portability.

2nd Dimension: Educational Impact

This dimension is concerned with assessing the benefits that could be gained by students when they use the system. Classically, these are done in pre- and post-tests, and this is carried on in this framework with more attention given to detail.

- Students grouped according to their mean grade in a quiz.
- Post-tests are used to compare one group with system only and another classroom only. A third group attends the classroom lecture with the class group and does a pre-test, and then uses the system before doing a post-test for comparison with the other two.
- Questions in the pre/post-tests must be mapped to each other to test the same types of knowledge, mainly consisting of declarative and procedural knowledge.
- The tests should best be attempted with students who were never exposed to this material previously to assess their learning rate.

3rd Dimension: Affective Measures

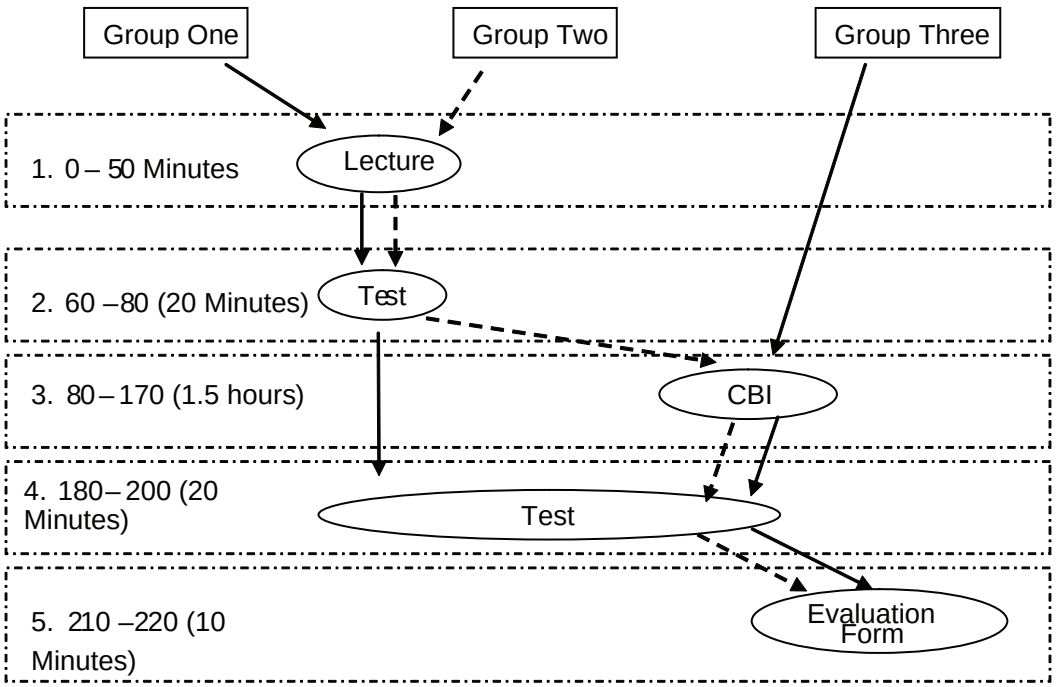
This dimension is mainly concerned with student opinions on the user friendliness of the system and allows them to express any shortcomings in the system. This could best be done through a survey where students are allowed to add any comments they wish freely and without restraints.

hypothesis" represents an investigation done between samples of the groups with the assumption that additional learning will not occur as a result of the treatment. In order to conduct a significance test, it is necessary to know the sampling distribution of F given that the significance level needed to investigate the null hypothesis. It must also be mentioned that the range of variation of averages is given by the standard deviation of the estimated means.

The ANOVA test did indeed show that there is a significant improvement in Group Two between the first test, which was taken after the lecture, and the second test, which was taken after using the system. However, this was not sufficient to be able to pinpoint the strengths of the system. Therefore, a more detailed analysis was done of student performance in the individual questions of test one and test two. Since the questions were mapped onto each other by design, it was easy to identify significant changes in student grades in a particular question type for students of group two who responded to similar questions before and after the use of the system. For example, a highly significant improvement with F=58 and $p<.000$ was observed in the question: "*Using an example, explain the stack concept and its possible use?*", which is an indication that the use of the system did strongly impact the student understanding of the concept of a "stack" in a functional manner.

Another point of view is to examine the scores by using the total average, which is 10.639, which can be approximated to 10.5, which can be

Figure 1. The evaluation procedure



used as a borderline. The rest of the scores can then be divided around this line. It was noticed that the average score of the third group was not high, yet 10 of the scores were above the borderline, while comparatively 6 scores were above it from the second group and only 6 of group one, which took the class only option. This shows the results of the third group, who took the CBI package alone and the second group, which had both the classroom lecture and the CBI package exposure to be close. It also underlines how much the second group improved their test results after taking the CBI and at the same time shows that the first group had not improved much only with the lecture learning. Alkhalifa (2001) tested in abstract logical context the effects of describing a state of a "*moving system*" versus describing it as a static system to find that this affects logical conclusions subjects eventually arrive at. This tutorial implemented those findings through this particular statistic, which revealed the effects of having animation in this multi-media system.

These results indicate that the use of the system may introduce a "limiting" effect that follows the initial introduction to the concepts (Albalooshi & Alkhalifa, 2002). Classroom lectures introduce students to the concepts, allowing them all the freedom to select all types of applications, which is in some ways overwhelming. The use of the system, on the other hand, produces a safe haven to test their ideas and strongly pursue the examples they can imagine, which helps them arrive at a solid procedural understanding of the concepts. It goes without saying that such a conclusion would have been impossible to make if the questions were not purposely set in the shown mapped fashion.

Additionally, students of groups two and three, who were exposed to the system, were asked to fill in an evaluation form composed of a series of questions as proposed by Caffarella (1987). They generally gave ratings of around 4 to 5 on a scale that went 0 to 6 with the highest for "The use of graphics, sound, and color contributes

to the student's achievement of the objectives," and "The user can control the sequence of topics within the CBI program." The lowest score was 3.559 for "The level of difficulty is appropriate for you". Therefore, it seems that the students in general enjoyed learning through the system, although they found the level of difficulty of the concepts presented challenging.

In addition to all this, three peer experts filled in evaluation forms to rate the system from an instructor's point of view and they gave the system an average rating of 5.33 on the same scale of 0 to 6.

FINE-GRAINED EVALUATION

The evaluation framework proposed here evaluates multi-media educational software at a finer level of detail than what was previously followed. The system architecture, for example, is evaluated in terms of the cognitive assumptions on which it relies. In many cases this dimension was overlooked, as in the studies that found animation and carefully organized images to be equivalent (Byrne et al., 1999; Lawrence et al., 1994) because they were not informed of the work done to show that a sequence of images are translated as animation (Freyd, 1987). In fact, there is a difference between the two in that animation presents subjects with a cognitive tool that reduces the cognitive load they endure while learning and allows them to concentrate on the transfer of information rather than assimilating the presented material into animation (Albalooshi & Alkhalifa, 2002).

The educational impact question must also be carried out at a finer level of detail than just to report total grades for comparison. Specific questions may be designed to probe key types of learning, as was shown in the presented case study.

FUTURE TRENDS

Only through a fine-grained analysis can the true features of the system be revealed. It is this type of information that allows designers to take advantage of the opportunities offered by an educational medium that offers to transform the learning experience into a joy for both students and educators alike. Consequently, cognitive science concepts represent themselves to future research as a viable means to comprehend the learning process. This would guide the evaluative process to focus on the points of strengths and weaknesses of each system rather than assess it as a whole unit.

CONCLUSION

A three-dimensional framework is presented as a means to evaluating multimedia educational software in order to resolve the shortcomings of the current evaluation techniques. It differs from the other, in that it seeks a more fine-grained analysis while being informed through cognitive science findings. The extra detailed review reveals specific strengths of the system that may have been otherwise concealed if the comparison was done in the traditional manner. In other words, this process focuses on students' cognitive interpretations of what they are presented with rather than assuming to know how they will approach the educational content presented to them.

REFERENCES

Al Balooshi, F., & Alkhalifa, E.M. (2002). Multimodality as a cognitive tool. In T. Okamoto & R. Hartley (Eds.), *Journal of International Forum of Educational Technology and Society, IEEE, Special Issues: Innovations in Learning Technology, 5*(4), 49-55.

Alkhalifa, E.M. (2001). Directional thought effect in the selection task. *Proceedings of the Third International Conference on Cognitive Science, ICCS 2001,* Beijing, China (pp. 171-176).

Alkhalifa, E.M., & Al Balooshi, F. (2003). A 3-dimensional framework for evaluating multimedia educational software. In F. Al Balooshi (Ed.), *Virtual education: Cases in learning & teaching technologies.* Hershey, PA: Idea Group Publishing.

Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of learning.* New York: McGraw-Hill.

Byrne, M.D., Catrambone, R., & Stasko, J.T. (1999). Evaluating animation as student aids in learning computer algorithms. *Computers and Education, 33*(4) 253-278.

Caffarella, E.P. (1987). Evaluating the new generation of computer-based instructional software. *Educational Technology, 27*(4), 19-24.

Freyd, J. (1987). Dynamic mental representations. *Psychological Review, 94*(4), 427-438.

Kinshuk, P.A., & Russell, D. (2000). A multi-institutional evaluation of intelligent tutoring tools in numeric disciplines. *Educational Technology & Society, 3*(4). *http://ifets.ieee.org/periodical/vol_4_2000/kinshuk.html*

Lawrence, W., Badre, A.N., & Stasko, J.T. (1994). *Empirically evaluating the use of animation to teach algorithms.* Technical Report GIT-GVU-94-07. Georgia Institute of Technology, Atlanta.

McKenna, S. (1995). Evaluating IMM: Issues for researchers. *Occasional Papers in Open and Distance Learning, 17.* Open Learning Institute, Charles Sturt University.

Pane, J.F., Corbett, A.T., & John, B.E. (1996, April). Assessing dynamics in computer-based instruction. *Proceedings of the 1996 ACM SIGCHI Conference on Human Factors in Computing Systems,* Vancouver, B.C. (pp.197-204).

Reiser, R.A., & Kegelmann, H.W. (1994). Evaluating instructional software: A review and critique of current methods. *Educational Technology, Research and Development, 42*(3), 63-69.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), *Curriculum evaluation.* Chicago: Rand-McNally.

Song, S.J., Cho, K.J., & Han, K.H. (2000). The effectiveness of cognitive load in multimedia learning. *Proceedings of the Korean Society for Cognitive Science* (pp. 93-98).

Song, S.J., Cho, K.J., & Han, K.H. (2001). Effects of presentation condition and content type on multimedia learning. *Proceedings of the Third International Conference on Cognitive Science, ICCS 2001,* Beijing, China (pp. 654-657).

Tam, M., Wedd, S., & McKerchar, M. (1997). Development and evaluation of a computer-based learning pilot project for teaching of holistic accounting concepts. *Australian Journal of Educational Technology, 13*(1), 54-67.

KEY TERMS

Cognition: The psychological result of perception, learning and reasoning.

Cognitive Load: The degree of cognitive processes required to accomplish a specific task.

Cognitive Science: The field of science concerned with cognition and includes parts of cognitive psychology, linguistics, computer science, cognitive neuroscience and philosophy of mind.

Cognitive Tool: A tool that reduces the cognitive load required by a specific task.

Declarative versus Procedural Knowledge: The verbalized form of knowledge versus the implemented form of knowledge.

Learning Style: This is the manner in which an individual acquires information.

Multimedia System: Any computer delivered electronic system that presents information through different media that may include text, sound, video computer graphics and animation.

ENDNOTE

[1] This term is defined in the Analysis of Results section.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 4, edited by M. Khosrow-Pour, pp. 2058-2062, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.16
Cognitive Functionality of Multimedia in Problem Solving

Robert Zheng
University of Utah, USA

ABSTRACT

Teaching problem solving can be a challenge to teachers. However, the challenge is oftentimes not due to a lack of skills on the part of learners but due to an inappropriate design of media through which the problem is presented. The findings of this study demonstrate that appropriately designed multimedia can improve learners' problem solving skills because of the cognitive functions media has in facilitating mental representation and information retrieval and maintenance, as well as reducing cognitive load during the problem solving process. Suggestions were made on how to apply interactive multimedia to teaching and learning.

INTRODUCTION

Problem solving skills as effective instructional strategies have been widely used in teaching and learning to enhance students' abilities to analyze, synthesize, and evaluate information during their learning processes (Hanley, 1987; Zheng & Zhou, 2006b). In the last five decades, researchers, teachers, and other educational practitioners have placed a heavy emphasis on this ability. For example, the movement of "discovery learning" (e.g., Bruner, 1961) was spawned, at least in part, by the perceived importance of fostering problem-solving skills. This emphasis on problem solving was not associated, however, with the knowledge of cognitive resources involved in problem solving. That is, it focused on the procedures of problem solving rather than investigating the relationship between the procedures of problem solving and cognitive resources that affect such procedures (Hanley, 1987; Sweller & Low, 1992). In the last 20 years, this state of affairs has begun to change with our knowledge of relevant mechanisms (e.g., working memory, cognitive load, etc.) increasing markedly. The introduction of multimedia, for instance, has reshaped our way of looking at how the information is processed in problem solving. Studies show that multimedia can improve learners' solving-problem skills, as appropriately designed multimedia tends to reduce learners'

cognitive load and increase cognitive resources during problem solving (Zheng & Smarkola, 2003; Zheng, Miller, Snelbecker, & Cohen, 2006a).

Although there are numerous studies on multimedia and cognition, little research has been done in the area of interactive multimedia and learner cognition, particularly in a situation where complex problems (e.g., multiple rule-based problems) are involved. Studies have shown that the interactive multimedia can be effective in simulating problem situations, thus providing a cognitive framework necessary for complex problem solving (Zheng et al., 2006a). Nonetheless, it is not clear whether such effectiveness is due to a difference in the design of multimedia or the cognitive functioning associated with such design. This chapter focuses on how interactive multimedia can facilitate learner problem solving by investigating:

1. The difference between interactive and non-interactive multimedia in complex problem solving
2. The difference between various forms of interactive multimedia and their impact on learners' information retrieval during complex problem solving

BACKGROUND

The Role of Media in Learning

More than 20 years ago, Clark (1983) challenged the educational community to reexamine its research design and results, claiming that most media research was confounded, and there were no significant differences among various media. Clark argued that the research should focus on method, aptitude, and task variables of instruction rather than media themselves. Kozma (1994) countered that research should be grounded in a theory that defines media in ways that are compatible and complementary with the cognitive and social processes by which knowledge is constructed. He urged researchers and practitioners to understand the dynamic relationships between media and learning, the reciprocal interaction between the learner's cognitive processes, and the unique characteristics of media. Reiser (1994) concurred with Kozma's view on media and pointed out that certain media attributes facilitate certain types of learning outcomes for particular types of learners.

Salomon (1979) in an earlier study observed that "something *within* the mediated stimulus … makes the presented information more comprehensible or better memorized by learners of particular characteristics" (p. 6). Based on the dual coding theory, Mayer and his associates (1997) investigated learners' information process in both multimedia and non-multimedia learning environments and concluded that appropriately designed multimedia can improve learning. According to the dual coding theory, learners learn better when information is presented through multiple sensory channels, rather than one channel only (Paivio, 1986). Studies (see Mayer, 1997; Mayer & Anderson, 1991; Mayer & Moreno, 2003) show that different presentation modes (i.e., words vs. pictures) and sensory modalities (i.e., audio vs. visual) may affect students' learning differently.

Working Memory and Cognitive Load

Research has found that input information such as auditory, visual, and kinesthetic information is processed through temporary storage before it is coded into the long-term memory (Baddeley & Logie, 1992; Logie, 1995). This temporary storage, also called working memory, comprises three major components: the phonological loop, visuo-spatial sketchpad, and central executive. The

phonological loop stores phonological information and prevents its decay by silently articulating its contents, thereby refreshing the information in a rehearsal loop. The central executive mechanism is believed to be related to cognitive activities such as reasoning and problem-solving. Finally, the visuo-spatial sketchpad (VSS) is believed to process and manipulate visuo-spatial images (Logie, 1995; Pearson & Logie, 2000). For example, the ability to mentally manipulate 3D images by rotating them in the mind is largely determined by the VSS function. The above components are closely related and interact among themselves in the process of problem solving (Bollaert, 2000; Mayer et al., 2003). Studies show that the working memory is very limited in both duration and capacity. Van Merrienboer and Sweller (2005) observed that the working memory stores about seven elements but normally operates on only two or three elements. When the working memory becomes overloaded with information, learning can be adversely affected (Marcus, Cooper, & Sweller, 1996; Sweller & Chandler, 1994).

Sweller et al. (1994) believe three types of cognitive load exist. They are *intrinsic load*, *extraneous* or *ineffective load*, and *germane* or *effective load*. The *intrinsic cognitive load* refers to the cognitive load that is induced by the structure and complexity of the instructional material. Usually, teachers or instructional designers can do little to influence the intrinsic cognitive load. The *extraneous cognitive load* refers to the cognitive load caused by the format and manner in which information is presented. For example, teachers may unwittingly increase learners' extraneous cognitive load by presenting materials that "require students to mentally integrate mutually referring, disparate sources of information" (Sweller et al., 1991, p.353). Finally, the *germane cognitive load* refers to cognitive load that is induced by learners' efforts to process and comprehend the material. Zheng et al. (2006a) asserted that learners' problem solving ability is correlated with the level of cognitive load. When learners become cognitively overloaded, that is, the working memory is filled with too much information, which leaves little room for thinking, their ability to solve problems can be impaired.

Recency Effect on Problem Solving

The ability to solve problems is not only affected by the amount of information held in the working memory, but also the status in which such critical information is maintained during problem solving process. While the process of problem solving draws on resources from both long-term and short-term memories, it is believed to rely heavily on the working memory for the working information in problem-solving (Baddeley et al., 1992; Logie, 1995). Capitani, Della Sala, Logie, and Spinner (as cited in Logie, 1995) conducted a study on learner information retrieval and found a high recall by subjects immediately after items have been presented. However if there was a filled delay before the recall was required, only the first few items on the list could be recalled. Logie (1995) described the former phenomenon as the recency effect and the latter as the primary effect. He believed that "the recency reflected the operation of a short-term or primary memory system" (p. 5), which was critical to the problem solving process.

It is essential to distinguish between cognitive load and recency effect because both involve cognitive resources in the working memory. The notion of cognitive load refers to the load or amount of information imposed on the learner's working memory while performing a particular task, whereas the recency effect refers to the most recent information that can be recalled in the working memory. The cognitive load study is focused on the working memory architecture and its limitations relating to the design of instruc-

tion (Paas, Tuovinen, Tabbers, & Gerven, 2003; Tabbers, Martens, & van Merrienboer, 2004; Van Merrienboer et al., 2005). The recency effect is, however, focused on the state of recalling and maintaining much needed information during the problem solving process. Zheng et al. (2006b) pointed out that one of the challenges to researchers and educators in applying multimedia to problem solving is to determine the optimal design where multimedia facilitates the recall and maintenance of critical information while keeping the working memory from becoming overloaded in the process of problem solving.

Problem Types and Cognitive Load

Problem types are related to thinking procedures in problem-solving processes (Delisle, 1997). For example, causal relationship problems require a linear thinking procedure that has a strong linear direction emphasizing the cause and effect whereas multiple rule-based problems involve simultaneously weighing several conditions/rules in mind in order to make a decision (Frye, Zelazo, & Palfai, 1995; Price & Yates, 1995). Zheng et al. (2006a) found that different types of problems may require different levels of working information in the problem solving process. For instance, multiple rule-based problem solving may require more working information than does causal relationship problem solving or single rule-based problem solving. Single rule-based problem solving, according to Frye et al. (1995), requires a straightforward deductive thinking such as applying the rule of card sorting to the action of sorting a deck of cards, whereas multiple rule-based problem solving involves a more complex, nonlinear thinking where the learner reaches a solution by engaging in a series of cognitive thinking activities such as analyzing, synthesizing, evaluating, and so forth while holding several conditions and rules in mind within a short time framework provided by the working memory (Johnson, Boyd, & Magnani, 1994; Price et al., 1995). Obviously, multiple rule-based problem solving is likely to increase intrinsic cognitive load more than the other two types of problem solving.

Spatial Ability and Multimedia Learning

Although use of multimedia for instruction can be beneficial, various learner characteristics have been found to mediate its effectiveness. One such characteristic is spatial ability. Pearson et al. (2000) noted that spatial manipulation is related to cognitive activities such as mental representation and synthesis. Logie (1995) studied the relationship among visuo-spatial ability, working memory, and problem solving and concluded that visuo-spatial working memory (VSWM) is involved in spatial-visual related problem solving and comprehension (also Pearson et al., 2000). Spatial strategies, either constructed mentally by the learner or presented as external instruction, are helpful in solving one- or two-dimensional problems including logical reasoning tasks and linear syllogisms, as well as inductive reasoning problems (Dupeyrat, 2000). Piburn and his colleagues (2005) investigated the role of visualization in computer based science education and found that learners' spatial ability is correlated with their achievement in learning supported by multimedia.

In short, learners' problem solving is affected by such factors as the amount of cognitive load in the working memory, the ability to retrieve critical information as well as maintain such information during problem solving, the types of problems that affect the information process in terms of cognitive load, and learner characteristics such as spatial ability. Characterized by multifold approaches of delivering information through multiple sensory channels, multimedia is believed to facilitate learners' problem solving.

However, existing literature is mainly focused on non-interactive multimedia studies. Little empirical research has been done to understand the cognitive functionality of interactive multimedia in problem solving.

THE STUDY

The study consisted of two sub-studies purported to find out whether there was (a) a difference between interactive and non-interactive multimedia in complex problem solving and (b) a difference between various forms of interactive multimedia and their impact on learners' information retrieval during complex problem solving. A description for each study follows.

Study 1

This study investigated whether there was a difference between interactive and non-interactive multimedia in complex problem solving and whether presentation modes were correlated with learners' spatial ability in problem solving. One hundred and fourteen students (80 undergraduate and 34 graduate students) were recruited from a large urban university in the northeastern region of the United States. Of 114 participants, 66 were females and 48 males. Participants varied in age from 19 to 58 years old, with a majority (74%) between the ages of 19 and 26.

The Instrumentation

Two instruments were used to measure (a) students' spatial ability and (b) problem solving skills. They include kit of factor-referenced cognitive tests: cube comparison and card rotation (Ekstrom, French, Harman, & Dermen, 1976) and interactive and non-interactive problem solving tasks.

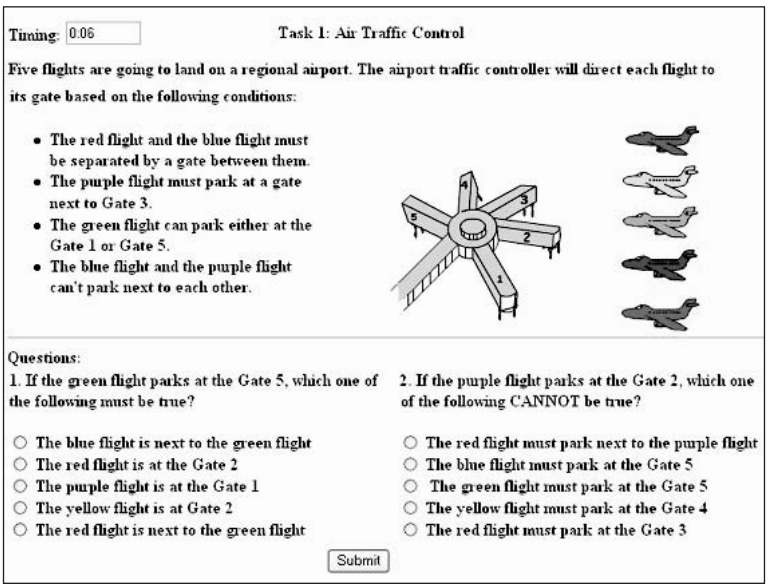
- **Kit of factor-referenced cognitive tests: Cube comparison and card rotation.** The cube comparison subtest has two parts. Parts one and two contain 21 items each, and subjects have three minutes to finish each part. Each item is a pair of cubes with different letters on the visual portions of the cube. Subjects must determine if the cubes are identical, presumably by mentally rotating one or both of the cubes. An individual's score on the cube comparison scale is calculated by subtracting the number of incorrect responses from the number of correct responses. Similar to the cube comparison subtest, the card rotation subtest has two parts. There are 10 items in each part and the subject has three minutes to finish each part. For each item, subjects are presented a target shape and eight different distractors similar to the target shape but at different rotated angles. For each of the eight distractors, subjects must determine if the distractor is the same as the target shape. An individual's total score on the card rotation scale is calculated by subtracting the number of incorrect responses from the number of correct responses on both parts.
- **Interactive and Non-interactive Problem Solving Tasks.** The problem solving tasks were developed by the author using Flash MX, Adobe Photoshop, and Microsoft Active Server Page (ASP). The five tasks were air traffic control, Tower of Hanoi, sailing boat, taking pictures, and office inspection. Each task consisted of two parts: (a) a problem presented with text format along with a visual presentation (either interactive or non-interactive), and (b) multiple choice questions. The tasks were multiple rule-based problems

that included a description of a problem situation and several mutually restricting conditions. The subject had to consider these conditions simultaneously before a solution could be reached. For example, Task 1, *air traffic control*, had a set of conditions that restricted the order and parking positions of airplanes. The subject had to consider all the conditions and then decide which flight would park at which gate without violating the conditions. Two multimedia versions of tasks were created: Interactive and non-interactive. In the interactive multimedia version, subjects were able to manipulate and move important components of the image (e.g., airplanes). In the non-interactive problem-solving version, subjects were given a static visual representation of each problem. For each problem, subjects were asked to answer two questions that measured the subject's problem solving skills. After completing the two questions, the subject clicked the submit button. A timer recorded the start and the end of the response time for each set of two questions (Figure 1).

Procedures and Conditions

Subjects were first given a spatial ability test and then categorized as low ability or high ability based on the test results. In the absence of normative data for this instrument, a median split was determined to be the most appropriate method for dividing the sample into low and high spatial ability. After the spatial ability groups were formed, subjects from the two ability levels were then randomly assigned to either the interactive or non-interactive problem solving condition. Thus, four groups were created: (a) high spatial ability/interactive, (b) high spatial ability/non-interactive, (c) low spatial ability/interactive, and (d) low spatial ability/non-interactive.

Figure 1. Sample of task 1: Air traffic control problem with questions



Results and Discussion

All data were performed with SPSS v.13. The means and standard deviations are provided in Table 1. The results of MANOVA indicated that the main effect for ability was not significant, $F(1,87) = 2.24$; $p = 1.38$. The main effect for type of multimedia presentation was significant, $F(1, 87) = 7.04$; $p < .05$. The interaction effect (ability x type of multimedia presentation) was not significant $F(1, 87) = 2.29$; $p = .134$. However, a graphic representation of the findings seems to indicate an interaction between the learner's spatial ability and interactive/non-interactive multimedia mode (Figure 2).

The findings revealed a difference between interactive and non-interactive multimedia in complex problem solving. Examination of the overall results leads to the impression that interactivity does seem to help both low and high ability learners, and that ability x presentation mode interactions may be detected with modifications of ability measures and problem solving criterion measures.

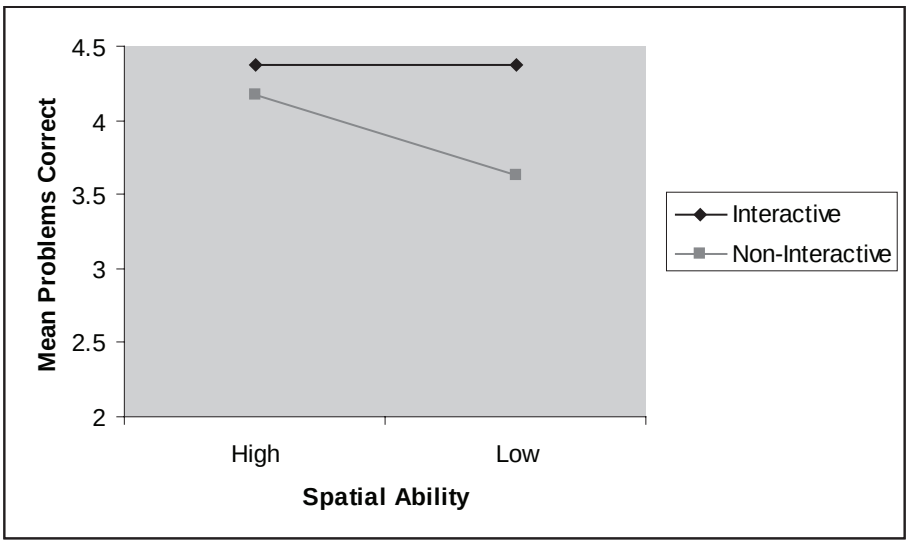
Study 2

Although the findings of study 1 indicated that interactive multimedia could facilitate learners' problem solving, it was not clear which mechanism within interactive multimedia was associated with the cognitive functioning in problem solving. Research has shown that differences in multimedia design such as simultaneous vs. successive presentation of verbal and visual information, can affect learners' information process during problem solving (Mayer, 2001; Mayer et al., 2003). It was thus hypothesized that the dif-

Table 1. Means and standard deviations for problem solving tasks

	Interactive	Non-Interactive	Margin Means
High Spatial Ability	4.38 (1.38)	3.96 (1.82)	4.17(1.60)
Low Spatial Ability	4.38 (1.66)	2.85 (2.00)	3.63 (1.97)
Margin Means	4.38 (1.50)	3.44 (1.97)	3.92 (1.79)

Figure 2. Spatial ability groups and multimedia modes



ferent design in interactive multimedia would affect learners' abilities to process information, which would in turn impact their way to solve complex problems.

Forty-five students were recruited from the same university as in study 1. Of 45 participants, 89% (n = 40) were undergraduate teacher education majors who received course credit for participating in the study and 11% (n = 5) were graduate students who volunteered to participate in this study. The average age of the participants was 21 with a range from 19 to 31.

The Instrumentation

Two instruments were used to measure students' spatial ability and problem solving skills. They include (a) the Guilford-Zimmerman Aptitude Survey Test 5: Spatial Orientation (Guilford-Zimmerman, 1956) and (b) synchronized and non-synchronized problem solving tasks.

- **Guilford-Zimmerman Aptitude Survey Test 5: Spatial orientation (Guilford-Zimmerman, 1956):** The Guilford-Zimmerman spatial orientation test requires participants to observe the position of the bow of a boat relative to the horizon and then to correctly identify the change in the boat's orientation by the change in horizon relative to the bow. The purpose of the Guilford-Zimmerman test is to assess participants' ability to mentally maneuver objects in terms of the special relationship, which is consistent with the reasoning tasks set up in this study: that is, the participant came up with a solution by maneuvering the figures and determining the spatial relationship between the figures based on the conditions and rules set in the tasks. The Guildford-Zimmerman spatial orientation test has reported a high reliability (alpha = .88) (Price & Eliot, 1975).

- **Synchronized and unsynchronized problem solving tasks:** Using the same problem set from study 1, the problem tasks in study 2 were modified to reflect a change in the task interface design. Two sets of problems were created: synchronized and unsynchronized interactive multimedia tasks. The synchronized interactive tasks included the problems and questions on the same page, whereas the unsynchronized multimedia tasks had two separate pages with one page containing the problems and the other the questions. The tasks contained five multiple rule-based problems. Each task included a description of the problem situation, several mutually restricting conditions, and two multiple-choice questions. If an error was made by the subject in any of these five problems, a point would be deducted from the total score. The total score equaled 10 points.

Procedures and Conditions

Using a blocked random sampling procedure, participants were first given the Guilford-Zimmerman Aptitude Survey Test 5: Spatial Orientation (Guilford & Zimmerman, 1956). The middle value of the distribution (median = 7.25) was determined and chosen as the cut score for defining subjects' high and low spatial ability. Subjects were then blocked into high and low spatial ability groups based on their spatial ability test score. Subjects were randomly drawn from each of the blocked groups (i.e., high and low spatial ability) to form two separate groups: the synchronized interactive multimedia group and the unsynchronized interactive multimedia group. Thus, two levels of interactive multimedia (synchronized vs. unsynchronized) were crossed with two levels of spatial ability (high vs. low) to form a 2 x 2 between subjects factorial design. In the synchronized

design, subjects were able to view the problems and questions on a single page whereas in the unsynchronized design subjects had to scroll between the pages to view the problems and questions. The independent variables included multimedia learning environments (synchronized vs. unsynchronized) and spatial ability (high vs. low). The dependent variables were test scores and efficiency scores.

Results and Discussion

The descriptive statistics for test scores are reported in Table 2. Significant differences were detected between the synchronized and the unsynchronized groups ($F_{(1, 44)} = 6.18$; $p < .05$) and between the high and the low spatial ability subjects in terms of test scores ($F_{(1, 44)} = 10.09$; $p < .01$).

The MANOVA analysis revealed that there was a main effect for multimedia groups (Wilks' Lambda = 5.64; $p < .05$) and spatial ability (Wilks' Lambda = 5.40; $p < .05$). An overall interaction between the multimedia groups and the spatial ability was detected (Wilks' Lambda = 6.48; $p < .01$) (Table 3).

Thus, with the first dependent variable, the results showed an impact of learners' performance in problem solving by the type of media employed. This was supported by the mean difference in test scores between the synchronized group (Mean $_{sync} = 5.43$) and the unsynchronized group (Mean $_{unsync} = 4.41$) (Table 2). The between-subjects tests showed a statistical significance between two groups ($F_{(1, 44)} = 6.19$; $p < .05$).

The second dependent variable of interest is efficiency, which is the number of problems solved correctly divided by the total time that it took the subject to complete the tasks. Results showed that both higher and lower spatial ability learners with the synchronized interactive multimedia obtained higher efficiency scores than their counterparts in the unsynchronized group (Table 4). There was an overall significant difference between synchronized and unsynchronized groups in efficiency scores ($F_{(1, 44)} = 11.99$, $p < .05$). Since the efficiency score indicates learners' ability to solve problems in terms of the ratio between total test scores and total response time (in seconds), it would be reasonable to assume that learners who solved more problems with less response time had more cognitive resources during

Table 2. Descriptive statistics for test scores

	Multimedia Group	Spatial Ability	Mean	Standard Deviation	N
Test Scores	Syncronized	High	6.17	1.26	12
		Low	4.64	1.50	11
		Total	5.43	1.56	23
	Unsynchronized	High	4.90	.56	10
		Low	4.00	1.47	12
		Total	4.41	1.22	22
	Total	High	5.59	1.18	22
		Low	4.30	1.49	23
		Total	4.93	1.48	45

Table 3. MANOVA analysis for multimedia groups and spatial ability

Effect		Value	F	Sig.	Partical Eta Squared
Group	Philai's Trace	.220	5.642	.007	.007
	Wilks' Lambda	.780	5.642	.007	.007
	Hotelling's Trace	.282	5.642	.007	.007
Spacial	Philai's Trace	.212	5.397	.008	.212
	Wilks' Lambda	.755	5.397	.008	.212
	Hotelling's Trace	.270	5.397	.008	.212
Group * Spacial	Philai's Trace	.245	6.480	.004	.245
	Wilks' Lambda	.755	6.480	.004	.245
	Hotelling's Trace	.324	6.480	.004	.245

the problem solving process. It would also imply that such learners were more effective in recalling and maintaining critical working information—a recency effect—and able to solve problems more efficiently in a short time framework provided by the working memory (Logie, 1995).

The study supported the hypothesis that the learner's ability to process information is affected by the design of *interactive* multimedia. Results indicated that synchronized interactive multimedia facilitated learners' abilities to solve problems due to a prompt recall and retrieval of working information and that unsynchronized interactive multimedia decreased learners' abilities to solve problems due to a filled delay phenomenon caused by scrolling back and forth between pages.

UNDERSTANDING COGNITIVE FUNCTIONS OF INTERACTIVE MULTIMEDIA

The previous findings support the notion that learners' abilities to solve complex problems can be influenced by the media with which the problems are presented. The findings also suggest that interactive multimedia facilitates cognitive functions in problem solving such as mental representation and information retrieval and maintenance.

Mental Representation

The dual coding theory posits that different forms of information, such as visual and auditory information, are registered through different sensory channels (Paivio, 1986). For example, the auditory information is registered through the auditory channel and the visual information is registered through the visual channel. The auditory and visual information then interact with each other within the working memory to form a mental representation of the external world (Mayer, 2001). Paivio (1986) believed that mental representation reflects the basic operation of comprehension and reasoning. When information is processed through multiple sensory channels, that is, through the channels of auditory, visual, and kinesthetic manipulation, the ability to formulate what has been perceived, namely, the mental representation of the external world, increases.

The results of our study indicated that there was a difference between interactive and non-interactive multimedia and that learners in the interactive multimedia group outperformed those in non-interactive multimedia group in problem solving (Table 1). It can be reasonably assumed that since learners with interactive multimedia were exposed to multiple sensory inputs including auditory, visual, and kinesthetic information, they were likely to form a more accurate and better mental representation of the external objects than did their counterparts with non-interactive multimedia. Researchers found that active engagement in learning such as manipulating the objects facilitates mental representation (Ploetzner, Bodemer, & Feuerlein, 2001). Clements (1999, cited from Reed, 2006) found that good manipulatives "provide control and flexibility, and assist the learner in making connections to cognitive and mathematical structures" (p. 95). Similarly, interactive multimedia enables learners to manipulate objects to simulate various situations for the possible solution, and therefore provides the kind of cognitive structures needed in problem solving.

Information Retrieval and Maintenance

Since the working memory is limited in both duration and capacity, complex problems may impose an excessive cognitive load on learners, thus affect their abilities to solve problems (Zheng et al., 2006b). According to Zheng et al. (2006a), complex problems, particularly multiple rule-based problems, have the characteristics of (a) simultaneously weighing several conditions/rules in mind before reaching a solution, and (b) consuming more cognitive resources than do single rule-based problems. They noted that it was the design of interactive multimedia rather than the interactive multimedia itself that affected the learner's ability to solve problems.

The results of this study indicated that synchronized interactive multimedia was more effective in solving complex problems than did unsynchronized interactive multimedia in terms of test scores and efficiency scores (Tables 2 and 4). This effectiveness is due to the cognitive functions that the synchronized mode of interactive multimedia displays in problem solving. Firstly, synchronized interactive multimedia tends to alleviate excessive cognitive load, thus improving efficiency in problem solving (Table 4). Unlike unsynchronized interactive multimedia, synchronized interactive multimedia provides an immediate access to the critical working information, and thus reduces the cognitive load for temporarily storing working information in problem solving. Secondly, synchronized interactive multimedia enables learners to retrieve and maintain much needed working information during problem solving. Unlike unsynchronized interactive multimedia that causes a filled delay in learners' information retrieval, synchronized interactive multimedia facilitates an immediate recall of critical information—a recency effect—during the problem solving process. Learners are able to retrieve and maintain much needed information while solving complex problems.

THE USE OF INTERACTIVE MULTIMEDIA IN TEACHING AND LEARNING

Multimedia, particularly interactive multimedia, has provided a richer environment for teaching and learning. Studies have shown that learning is effective for interactive visual images (Marschark & Hunt, 1989, Reed, 2006; Zheng et al., 2006a). The findings of this study have revealed that inter-

Table 4. Descriptive statistics for efficiency scores

	Multimedia Group	Spatial Ability	Mean	Standard Deviation	N
Efficiency Scores	Syncronized	High	.39	.08	12
		Low	.26	.09	11
		Total	.33	.11	23
	Unsynchronized	High	.26	.03	10
		Low	.22	.07	12
		Total	.24	.06	22
	Total	High	.33	.11	22
		Low	.24	.06	23
		Total	.29	.09	45

active multimedia supports mental representation. Further, synchronized interactive multimedia supports the cognitive functioning in problem solving by reducing cognitive load and facilitating information retrieval and maintenance. Since learners' abilities to solve problems are closely related to their perceptions, the ability to form a mental representation of the external world is critical to the problem solving process. Since the ability to *effectively* solve problems is affected by the availability of cognitive resources, the way to increase such cognitive resources by reducing the cognitive load is essential in facilitating an environment for successful problem solving. Based on the discussions above, we would like to advance several suggestions regarding the use of interactive multimedia in teaching and learning:

1. It is suggested that the use of interactive multimedia should be based on an understanding of the cognitive functions of interactive multimedia. That is, an understanding of the relationship between the attributes of interactive multimedia and the cognitive functions in learning.
2. It is suggested that the use of interactive multimedia should focus on how to optimally maximize the effects of input information such as auditory, visual, and kinesthetic information in teaching and learning, especially the impact of such input information on cognitive learning such as learners' mental representation of external objects.
3. It is suggested that the use of interactive multimedia should take into consideration the design factor as well as the impact of the design on learners' cognition, particularly learners' ability to reduce cognitive load and retrieve and maintain critical information in learning.

CONCLUSION

Teaching problem solving can be challenging to teachers. Oftentimes, such a challenge is not due to learners' lack of skills to solve problems but because of inappropriately designed media that may cause cognitive overload, which leaves little cognitive resource for learners to work on the problems. Studies have shown that the learner's ability to solve problems can be mediated by an inadequate cognitive resource in the working memory resulting from a cognitive overload (Mar-

cus et al., 1996; Mayer, 2001; Van Merrienboer et al., 2005). The findings of this study demonstrated that appropriately designed media such as interactive multimedia can improve learners' abilities to solve complex problems. Moreover, appropriately designed media can be complementary to individual differences. For example, interactive multimedia can improve the performance of low spatial ability learners in problem solving. One of the important findings of this study was that synchronized interactive multimedia was effective in reducing cognitive load and facilitating information retrieval and maintenance during the problem solving process.

The study is significant in that it investigated, for the first time, the cognitive functions of interactive multimedia, especially synchronized interactive multimedia in complex problem solving. It probed into the attributes of interactive multimedia and their relations with the cognitive functions of problem solving. Future study should examine the relationship between various interactive multimedia designs and their cognitive functions in learning. It is suggested that future study should expand to include what Snow (1996) described as conative domain and examine learners' self-efficacy in complex problem solving in an interactive multimedia environment.

REFERENCES

Baddeley, A. D. (1999). *Essentials of human memory.* Hove, UK: Psychology Press.

Baddeley, A. D., & Logie, R. H. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory imagery* (pp. 179-197). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bollaert, M. (2000). A connectionist model of the processes involved in generating and exploring visual mental images. In S. O. Nuallian (Ed.), *Spatial cognition: Foundations and applications.* (pp. 329-346). Amsterdam/Philadelphia: Joint Benjamins Publishing.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct mearsurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Bruner, J. (1961). The act of discovery. *Harvard Educational Review, 31*(1), 21-32.

Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Delisle, R. (1997). *How to use problem-based learning in the classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

Dupeyrat, M. G. D. (2000). Spatial strategies in reasoning. In G. D'Ydewalle, D. De Vooght, W. Schaeken, & Vandierendonck, A. (Eds.), *Deductive reasoning and strategies* (pp. 153-175). Mahwah, NJ.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Services.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483-527.

Guilford, J. P., & Zimmerman, W. S. (1956). *Guilford-Zimmerman aptitude survey.* Beverly Hills, CA: Sheridan.

Hanley, G. L. (1987). *The origin of information and its effects on problem solving.* Paper presented at the Annual Meeting of the American Psychological Association. New York.

Johnson, J. T., Boyd, K. R., & Magnani, P. S. (1994). Causal reasoning in the attribution of rare

and common events. *Journal of Personality and Social Psychology, 66*(2), 229-242.

Kozma, R. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research & Development, 42*(2), 7-19.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, UK: Lawrence Erlbaum Associates.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*(1), 49-63.

Marschark, M., & Hunt, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Human Learning and Memory, 15,* 710-720.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge, UK: Cambridge University Press.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E., & Anderson, R. (1991). Animations and narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83,* 484-490.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Paas, F., Tuovinen, J. E., Tabbers, H., & Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Pearson, D. G., & Logie, R. H. (2000). Working memory and mental synthesis: A dual task approach. In S. O. Nuallian (Ed.), *Spatial cognition: Foundations and applications.* (pp. 347-359). Amsterdam/Philadelphia: Joint Benjamins Publishing.

Piburn, M. D., Reynolds, S. J., McAuliffe, C., Leedy, D. E., Birk, J. P., & Johnson, J. K. (2005). The role of visualization in learning from computer-based images. *International Journal of Science Education, 27*(5), 513-527.

Ploetzner, R., Bodemer, D., & Feuerlein, I. (2001). *Facilitating the mental integration of multiple sources of information in multimedia learning environments***.** Paper presented at ED-Media 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Tampere, Finland.

Price, L., & Eliot, J. (1975). Convergent and discriminant validities of two sets of measures of spatial orientation and visualization. *Educational and Psychological Measurement, 35*(4), 975-977.

Price, P. C., & Yates, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology, 21*(6), 1639-1655.

Reed, S. K. (2006). Cognitive architectures for multimedia learning. *Educational Psychologist, 41*(2), 87-98.

Reiser, R. (1994). Clark's invitation to the dance: An instructional designer's response. *Educational Technology Research & Development, 42*(2), 45-48.

Salomon, G. (1979). *Interaction of media, cognition, and learning.* San Francisco: Jossey-Bass.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law, 2*(3/4), 536-560.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*(4), 351-362.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Tabbers, H. K., Martens, R. L., & van Merrienboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71-81.

Sweller, J., & Low, R. (1992). Some cognitive factors relevant to mathematics instruction. *Mathematics Education Research Journal, 4*(1), 83-94.

Van Merrienboer, J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177.

Zheng, R., & Smarkola, C. (2003) Multimedia learning environments for early readers. *Academic Exchange Quarterly, 7*(4), 229-32.

Zheng, R., & Zhou, B. (2006b). Recency effect on problem solving in interactive multimedia learning. *Journal of Educational Technology and Society, 9*(2), 107-118.

Zheng, R., Miller, S., & Snelbecker, G. (2005). *Which works: Media or methods?* (MAR*TEC Techno-Brief, No. 149). Temple University, PA: The Mid-Atlantic Regional Technology in Education Consortium.

Zheng, R., Miller, S., Snelbecker, G., & Cohen, I. (2006a). Use of multimedia for problem-solving tasks. *Journal of Technology, Instruction, Cognition, and Learning, 3*(1-2), 135-143.

KEY TERMS

Cognitive Load: According to cognitive load theory (CLT), three types of cognitive load exist: *intrinsic load*, *extraneous* or *ineffective load*, and *germane* or *effective load*. The *intrinsic cognitive load* refers to cognitive load that is induced by the structure and complexity of the instructional material. Usually, teachers or instructional designers can do little to influence the intrinsic cognitive load. The *extraneous cognitive load* refers to the cognitive load caused by the format and manner in which information is presented. For example, teachers may unwittingly increase learner's extraneous cognitive load by presenting materials that "require students to mentally integrate mutually referring, disparate sources of information" (Sweller et al., 1991, p. 353). Finally, the *germane cognitive load* refers to cognitive load that is induced by learners' efforts to process and comprehend the material. The goal of CLT is to increase this type of cognitive load so that the learner can have more cognitive resources available to solve problems (Brunken, Plass, & Leutner, 2003; Marcus et al., 1996).

The Dual Coding Theory: The dual coding theory describes the role of sensory inputs in information processing. According to the dual coding theory, different forms of information such as verbal, visual, and auditory information are registered through different channels (Paivio, 1986). For example, the auditory information is registered through the auditory channel and the visual information is registered through the visual channel. The auditory and visual information then interact with each other within the working memory to form a mental representation of the external world (Mayer, 2001). Paivio (1986) argued that learners learn better when information is presented through multiple sensory channels, rather than one channel only. Studies (see Mayer, 1997; Mayer et al., 1991; Mayer et al., 2003) show

that different presentation modes (i.e., words vs. pictures) and sensory modalities (i.e., audio vs. visual) may affect students' learning differently.

Interactive Multimedia: Interactive multimedia refers to the use of several media in learning where learners are able to process information through multiple sensory channels including auditory, visual, and kinesthetic manipulation. The advantages of interactive multimedia in learning include visualizing abstract and concrete ideas by creating images, diagrams, or animations, reducing cognitive load in learning, facilitating mental representation of external objects, and improving cognitive learning for low spatial ability learners.

Multiple Rule-Based Problems: Multiple rule-based problems refer to the type of problems that consist of problems, rules and conditions that are mutually restricting. The learner is to find an optimal solution by weighing the conditions and rules and at the same time make a decision that would meet the conditions or rules without conflicting each other. The multiple rule-based problem involves a complex, nonlinear thinking process where the learner reaches a solution by engaging in a series of cognitive thinking activities such as analyzing, synthesizing, and evaluating the information while holding the conditions and rules in mind within a short time framework provided by the working memory. Thus, multiple rule-based problem solving may require more working information than does causal relationship problem solving or single rule-based problem solving. The multiple rule-based problem solving is likely to increase intrinsic cognitive load more than other two types of problem solving.

Single Rule-Based Problems: Single rule-based problems refer to a type of problem that emphasizes causal relationship among entities. The single rule-based problem solving, according to Frye et al. (1995), focuses primarily on the cause and effect of events and requires straightforward deductive thinking such as applying the rule of card sorting to the action of sorting a deck of cards. Studies show that single rule-based problems may require less working information within the working memory and impose less cognitive load on learners than do multiple rule-based problems.

Visualization: Visualization refers to techniques used to communicate both abstract and concrete ideas by creating images, diagrams, or animations.Visualization has been defined as an important indicator of measuring learners' spacial ability. Its applications have expanded into science, engineering, education, medicine, etc. Typical applications of visualization include computer graphics, interactive multimedia, animations, and so forth.

Working Memory: Working memory is a theoretical framework that refers to the structures and processes used for temporarily storing and manipulating information. According to Baddeley and Hitch (1974), the working memory consists of two "slave systems" responsible for short-term maintenance of information, and a "central executive" responsible for the supervision of information integration and for coordinating the slave systems. One slave system, the articulatory loop, stores phonological information and prevents its decay by silently articulating its contents, thereby refreshing the information in a rehearsal loop. The other slave system, the visuo-spatial sketch pad, stores visual and spatial information. It can be used, for example, for constructing and manipulating visual images, and for the representation of mental maps. The sketch pad can be further broken down into a visual subsystem (dealing with, for instance, shape, color, and texture), and a spatial subsystem (dealing with location). The central executive system is, among other things, responsible for directing attention to relevant in-

formation, suppressing irrelevant information and inappropriate actions, and coordinating cognitive processes when more than one task must be done at the same time. Studies show that the working memory is very limited in both duration and capacity. The working memory typically stores about seven elements but normally operates on only two or three elements.

This work was previously published in Handbook of Research on Instructional Systems and Technology, edited by T. T. Kidd and H. Song, pp. 232-248, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.17
Multimedia, Information Compexity and Cognitive Processing

Hayward P. Andres
Portland State University, USA

ABSTRACT

Organizations are faced with increasing costs needed to train employees in today's high technology environment. Educators are also striving to develop new training and teaching methods that will yield optimal learning transfer and complex skill acquisition. This study suggests that trainee/learner cognitive processing capacity, information presentation format and complexity, and multimedia technology should be leveraged in order to minimize training duration and costs and maximize knowledge transfer. It presents a causal model of how multimedia and information complexity interact to influence sustained attention, mental effort and information processing quality, all of which subsequently impact comprehension and learner confidence and satisfaction outcomes. Subjects read a text script, viewed an acetate overhead slide presentation containing text-with-graphics, or viewed a multimedia presentation depicting the greenhouse effect (low complexity) or photocopier operation (high complexity). Causal path analysis results indicated that presentation media (or format) had a direct impact on sustained attention, mental effort, information processing quality, comprehension, and learner confidence and satisfaction. Information complexity had direct effects on sustained attention, mental effort and information processing quality. Finally, comprehension and learner confidence and satisfaction were both influenced through an intervening sequence of sustained attention, mental effort and information processing quality.

INTRODUCTION

During information presentation, the target audience must construct a mental representation of situations or scenarios conveyed by the verbal content and images contained in the presentation. Cognitive psychologists refer to these representations as situation models (Friedman & Miyake,

2000). During situation model construction, increases in the number of alternative order of events, number of interconnections among objects and events, and factors that give rise to specific events will lead to a decline in the accuracy and capacity in cognitive processing utilized to construct a situation model (Zwaan, Magliano, & Graesser, 1995).

During multimedia presentation, subjects are presented with information in verbal and pictorial form, and both the verbal and visual processing channels of memory are used to translate the information into the appropriate situation model (Hegarty, Narayanan, & Freitas, 2002; Mayer & Moreno, 2002). In instructional settings, animation and other types of graphics that depict the behavior of various phenomena such as meteorology, physics, or chemistry have been used to reduce information complexity, augment cognitive processing, and facilitate comprehension (Moreno & Mayer, 2002; Rieber, 1991). Multimedia can also reduce the perceived equivocality of a low-analyzable decision-making task (Lim & Benbasat, 2000) and promote computer self-efficacy that leads to increased performance in computer-based training situations (Christoph, Schoenfeld & Tansky, 1998).

The goal of this study is to investigate the impact of multimedia information representation on cognitive processing activities (e.g., information encoding, situation model construction, and comprehension) typical to problem solving, training, and decision-making contexts. A capacity theory of comprehension (Just & Carpenter, 1992), dual processing theory of working memory (Mayer & Moreno, 2002; Paivio, 1986), theory of attentional inertia (Burns & Anderson, 1993), and the PASS (Planning, Attention, Simultaneous, and Successive) cognitive processing theory (Naglieri & Das, 1997) are used to provide a framework for this investigation.

The following section presents a review of empirical research on information presentation mode (i.e., format), information complexity, and cognitive processing. Next, using relevant research findings, a causal path model that presents hypothesized linkages among information presentation mode, information complexity, sustained attention and mental effort, information processing quality, comprehension, and learner confidence and satisfaction is presented. This is followed by a discussion of the findings, implications of results, and suggestions for future research on assessing multimedia-based information presentation on cognitive processing in learning, training, and decision-making settings.

BACKGROUND AND THEORETICAL FRAMEWORK

Information Presentation Media

Visual imagery depicts spatial arrangement, relative size, physical appearance, and the configuration of sub-components. Levin, Anglin, and Carney (1987) noted that pictures have the following effects: (1) minimize explanatory content by summarizing distinctive features or procedures; (2) facilitate interpretation by clarifying abstract concepts; (3) facilitate comprehension by eliminating the need to translate text into imagery; and (4) facilitate long-term memory by creating a memorable mnemonic. Mayer and Moreno (1998) noted that verbal (text or auditory) and visual information are each processed through distinct cognitive processing channels (i.e., verbal and visual) that compliment each other.

Multimedia utilizes computer and audio-visual technology to present information verbally (text or auditory), as static pictures or diagrams, and as animated graphics or video (Kozma, 1991). Attentional inertia (i.e., sustained attention and applied mental effort) results when a presentation medium induces perceptual arousal that sustains attention to the medium, and when learner confidence and satisfaction is promoted through enhanced cognitive processing (Burns & Anderson, 1993).

Information Complexity

Inferential complexity associated with information is a function of the number of causal links in a chain of actions, physical states, or mental states. Causal links between units of information are typically defined through the use of temporal (e.g., before, and then, after), causal (e.g., which caused, which enabled, because, if-then), or intentional (e.g., in order that, so that, to allow) connectives (Millis, Golding, & Baker, 1995). As the number of causal links needed to convey an idea or concept increases (i.e., causal chain length), working memory becomes overloaded because it is limited in the number of related ideas that can be simultaneously stored and processed (Halford, Wilson & Philips, 1998). Halford et al. (1998) noted that when working memory capacity is exceeded, subjects begin to condense the situation model to reduce cognitive load, but at the expense of making some relational information, that may be needed in subsequent processing, inaccessible. Information complexity is also a function of the extent to which one clause is related to more than one other clause–connective span (Millis, Graesser, & Haberlandt, 1993).

Cognitive Processing

The dual-coding theory of working memory suggests that information encoding and processing can take place in a verbal working memory and/or in a visual working memory workspace (Paivio, 1986). Comprehension is enhanced as a result of a reduction in cognitive load because the visual working memory workspace immediately encodes spatial information and does not require any translation of verbal information into imagery (Mayer & Moreno, 2002; Moreno & Mayer 2002). According to PASS cognitive processing theory, effective cognitive processing is controlled by an executive function responsible for selective attention, sustained attention, attentional switching, and mental effort, while encoding incoming verbal and/or spatial information and constructing the situation model (Naglieri & Das, 1997). Recent studies have suggested that greater human information processing and comprehension outcomes in complex sequential cognitive tasks are associated with increased mental effort or focused and directed concentration (Rende, Ramsberger & Miyake, 2002; Washburn & Putney, 2001; van Merrienboer, Schuurman, de Croock, & Paas, 2002).

Comprehension

The main contention of the capacity theory of comprehension is that cognitive capacity (e.g., short-term and long-term memory) facilitates or constrains computational (i.e., causal analysis of noun-verb clauses) and storage demands imposed in the construction of situation models (Just & Carpenter, 1992). Gordon, Hendrick, and Levine (2002) observed lower comprehension when subjects attempted to remember a short set of words while reading syntactically complex sentences. McElree (2000) suggested that information is maintained in working memory via a content-addressable mechanism and this information is periodically accessed to fill in gaps existing in the current situation model. Some studies have also shown that comprehension is also enhanced when information presentation takes place in multi-modal form (i.e., verbal, images, animation) thereby enhancing information processing by making use of the additive and synergistic properties of the verbal and visual working memory systems (Mayer & Chandler, 2001; Mayer & Moreno, 2002; Park, 1998).

Learning Satisfaction

Learning satisfaction has been described as a sense of accomplishment that learners feel at the conclusion of a learning event when the learning environment facilitated information processing that led to successful comprehension outcomes

(Keller, 1987; Song & Keller, 2001). In Keller's (1987) ARCS Model of Motivational Design, effective learning contexts exhibit four essential conditions – attention, relevance, confidence, and satisfaction. Learning satisfaction is highest when the information presentation arouses and maintains attention, curiosity and interest throughout the entire duration of the presentation. Further, attention is maintained when the instructional content is coherent and conveys relevant importance and meaningfulness to the learner. Finally, learning satisfaction arises from learner confidence that understanding of the content has been achieved and there is a positive correlation between learner mental effort and the extent of learning achievement (Keller, 1987; Small & Gluck, 1994). Self-reported learning satisfaction has also been associated with evaluations of the learning time and effort efficiency (Cole, 1992), willingness to learn more about the topic (Maki, Maki, Patterson & Whittaker, 2000), and perceived ease at learning (Hackman & Walker, 1990).

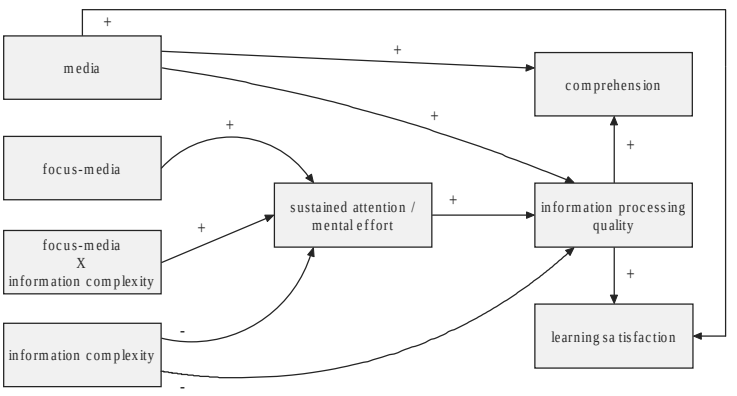
RESEARCH HYPOTHESES DEVELOPMENT

Figure 1 below graphically depicts the theoretical framework used in the formulation of the research hypotheses. In this study, the causal model presented suggests that information presentation media (i.e., text-only, graphics-and-text, and multimedia) and information complexity will each directly impact information processing quality (i.e., understandable, informative, interesting) and will interact to exert influences on the amount of sustained attention and mental effort applied in cognitive processing (i.e., situation model construction) of the information presented. Sustained attention and mental effort would then have an intervening influence on information processing quality. Finally, information processing quality would then impact comprehension and learning satisfaction outcomes.

Mental Effort

Cognitive processing of text-based information imposes a required involuntary demand for applied mental effort and sustained attention because sentences must be parsed to identify and extract

Figure 1: Impact of media and information complexity on information processing and comprehension outcomes.



noun-verb linkages representing concepts or ideas. Mental effort is then further extended when these concepts are placed into a situation model that is then mentally simulated for comprehension, verification, and self-correction. Alternatively, Burns and Anderson (1993) noted that increased sustained attention was associated with presentation media with the greatest sensory stimulation and appeal (i.e., dynamic visual and auditory stimuli). Sensory appeal and sustained attention were associated with the subjects' anticipation of forthcoming content. Therefore,

H1: *Text-based and multimedia information presentations, when compared to graphics with text information presentations, will lead to greater applied sustained attention and mental effort.*

Information complexity has also been associated with sustained attention and applied mental effort (Burns & Anderson, 1993). Content that is very difficult to encode and place into an appropriate situation model lowers sustained attention because the complex content is lower in intellective appeal, and the subject determines that applying continued attention and mental effort to forthcoming information will not be beneficial. Further, sustained attention and mental effort are inhibited when the subject's ability to understand the information presented is lowered by complex information (Christoph et al., 1998). Therefore,

H2: *As information content increases in complexity, subjects will expend less sustained attention and mental effort.*

Information presentation media and information complexity can interact to influence applied mental effort and sustained attention. As information complexity increased or the task required greater imagery or visualization, visual presentation reduced cognitive load through immediate encoding (Burns & Anderson, 1993). In addition, the sensory features of multimedia such as visual complexity, movement, and sound effects created an "audio-visual momentum" that led to perceptual arousal, which subsequently induced sustained attention and mental effort (Burns & Anderson, 1993). The "audio-visual momentum" reduces the tendency to withdraw attention and mental effort typically associated with the presentation of complex information. Therefore,

H3: *As information complexity increases (decreases), text-only and multimedia information presentations will lead to greater (lesser) sustained attention and mental effort when compared to graphics-with-text information presentations.*

Information Processing Quality

Mayer and Moreno (1998) noted that when using multimedia, subjects were able to integrate words and pictorial animation of lightening formation more easily when the words supplementing the animation were presented auditorily as opposed to written text. The encoding of the auditory/animation information presentation was more efficient because the subjects were able to maintain complete visual attention on the animation while simultaneously attending to the narration content. In contrast, the animation/text group experienced diminished cognitive processing due to a "split-attention" effect (i.e., required alternating shifts of visual focus between text processing and visual imagery processing). Therefore,

H4: *As information presentation format increases in modality (i.e., from text-only to graphics-with-text to multimedia), subjects will experience greater information processing quality.*

Rieber's (1991) study comparing static graphics versus animated graphics observed better knowledge extraction with animated graphics and, when given a choice to return to either presentation format, subjects overwhelmingly preferred the animated graphics presentation. The subjects in

Rieber's (1991) study exhibited attentional inertia (i.e., sustained attention and applied mental effort) when viewing the animated graphic presentation. The subjects found the multimedia presentation intellectually appealing, easier to understand, and more interesting. In addition, increased information processing quality was evidenced by the fact that the subjects viewing animation were able to acquire incidental knowledge when making inferences from imagery than with text-based propositions constructed from static graphics. Therefore,

H5: *An increase in sustained attention and mental effort will lead to greater information processing quality.*

Information complexity can directly influence information processing quality through an imposed increase in cognitive load. Long causal chain length can create difficulty in acquiring local coherence within a situation model and extensive hierarchical connections and high connective spans among related concepts can create difficulty in acquiring global coherence. Friedman and Miyake (2000) noted that casual connectives with long causal chains or high connective span temporarily suspended purging of the working memory in attempts to construct a complete and coherent situation model. Failure to purge working memory caused working memory overload. Consequently,

H6: *As information complexity increases, subjects will experience lower information processing quality.*

Comprehension and Learning Satisfaction

Information processing quality mediates comprehension because multimedia can lead to a reduction of complexity and increased cognitive processing capacity by targeting both verbal and visual working memory channels, thereby making the information presented more understandable, interesting, and informative. Alternatively, because of its linear structure, a text presentation that is structurally equivalent (i.e., equality in the setting, characters and/or objects, conceptual content, and episodic structure) to a static image with supporting verbal information or multimedia presentation will require more cognitive processing capacity to process (Mayer & Moreno, 1998, 2002). This suggests the following hypothesis.

H7: *As information processing quality increases, subjects will experience greater comprehension.*

In the comparison of text, narration, and graphics with text, Mousavi, Low, and Sweller (1995) found that a combined text and graphics presentation reduced the cognitive load in solving geometry problems resulting in superior learning. The graphics depicting the geometric problem space clarified the abstract equation representation; this enabled timely solution to the problem. Cassady (1998) noted that, compared to traditional lecture presentations, students found multimedia-based presentations to be superior in the following areas: (1) flow, organization, and understandability of the information presented; (2) ease in following the presentation; (3) ability to pay attention to the presentation; and (4) interest in the information presented. Therefore,

H8: *As information presentation format increases in modality (i.e., from text-only to graphics-with-text to multimedia), subjects will experience greater comprehension.*

Towler and Dipboye (2001) observed superior training outcomes when the training presentation was characterized as higher in coherent organization and trainer expressiveness. Effective training presentations provided clarifying and elaborative content and commanded sustained

attention through appropriate vocal intonations that acted as cues that made the presentation easy to follow. The trainees also displayed greater positive affect, information recall and comprehension, and perceptions that they could apply what they learned to actual task execution and problem solving. After 12 weeks of study and practice in an introductory programming class, Ramalingam and Wiedenbeck (1998) observed an increase in subject-perceived computer-programming learning achievement. Consequently,

H9: *As information processing quality increases, subjects will experience greater learning satisfaction.*

H10: *As information presentation format increases in modality (i.e., from text-only to graphics-with-text to multimedia), subjects will experience greater learning satisfaction.*

RESEARCH METHODOLOGY

Participants

Seventy-eight male and female undergraduate students voluntarily participated in the experiment. Each subject was eligible to participate in a lottery of five drawings of fifty dollars. The average age was 25 years (*SD = 5.39*).

Materials

Two text passages, *Photocopier Operation* (high information complexity) and *GreenHouse Effect* (low information complexity), with relatively equal readability scores were utilized. Every attempt was made to keep the following factors as constant as possible between the passages: text length, Flesch Reading Ease, and Flesch-Kincaid Reading Grade Level (Wagenaar, Schreuder, & Wijlhuizen, 1987). The Flesch Reading Ease statistic is calculated by multiplying the average sentence length by 1.015, multiplying the number of syllables per 100 words by .846, and then the sum of these products is subtracted from 206.835. The Flesch-Kincaid Reading Grade level statistic is calculated by adding the average sentence length to the percentage of long words (i.e., more than two syllables), and the sum is then multiplied with 0.4. Table 1 details specific readability scores.

For complexity, the passages differed in causal chain lengths and the extent to which one clause was causally related to other clauses throughout other parts of the scenario (i.e., connective span). Connective span was computed as the number of causal relationships each idea unit had with other idea units within the same sentence and across sentences throughout the text. The Greenhouse Effect and Photocopier Operation, respectively, had 9 and 18 idea units with a connective span of three and 10 and 29 sentences with direct contigu-

Table 1: Passage Readability Measures

Readability Measure	**Photocopier Operation**	**GreenHouse Effect**
Paragraphs	14	12
Sentences (words/sentence)	8 6 (19.6) 8	5 (16.8)
Flesch Reading Ease	44.8 4	0.7
Flesch-Kincaid Grade Level	11.8	11.7

ous causal links greater than one. The connective span and direct causal links values indicated that the Photocopier Operation passage was higher in information complexity than the Greenhouse Effect passage.

The overhead slides and multimedia presentations were both equivalent in overall content (i.e., idea units expressed) and episodic structure (i.e., sequence of events) to the text passages (Baggett, 1979). The overhead-slides presentation utilized static diagrams that illustrated components involved in photocopier operation or elements involved in greenhouse gas generation. The multimedia presentation animated the elements contained in the static diagrams used in the overhead presentation.

Procedure

Path analysis was the chosen as the data analysis procedure because analysis of variance is not particularly well suited for testing direct, mediating and intervening relationships among a set of variables. Each subject was randomly assigned to one of the media (i.e., text-only, text with graphics, or multimedia) and complexity conditions (i.e., high or low information complexity). The overhead slides sessions presented the same conceptual content using static overheads composed of graphs and pictures supplemented with text. Subjects in the computer-based multimedia presentation sessions watched a series of PowerPoint slides that contained animated graphs and images along with supporting text. All sessions (text-only, text-with-graphics overhead slides, and multimedia) were executed for a duration of 12 minutes. After each session, the subjects completed a Likert-type questionnaire that elicited perceptions of the amount of sustained attention/mental effort, information processing quality, and learning satisfaction. Following this, they were administered a timed comprehension test composed of short essay questions.

Measures

Information Presentation Media. Presentation media (i.e., text-only, text-with-graphics, or multimedia) was coded using two contrast code variables (Pedhazur, 1997). The first contrast code variable, *media*, contrasted group mean outcomes of each of the three presentation modes (i.e., text-only, text-with-graphics, and multimedia). The second contrast code variable, *focus-media*, contrasted the group means of text-only and multimedia combined against the mean of the text-with-graphics group.

Information Complexity. Information complexity was operationalized as a function of connective span (i.e., the number of causal relationships each idea unit had with other idea units within the same sentence and across sentences throughout the text) and causal chain length (i.e., number of contiguous causal connections). The Photocopier Operation content was greater in information complexity than the Greenhouse Effect content. Information complexity was contrast coded. Grade Reading Level scores for both passages were similar in value in order to restrict complexity to connective span and causal chain length (see Table 1) and to minimize confounding from differences in domain content.

Sustained Attention/Mental Effort. The sustained attention/mental effort scale (Likert-type) was developed using Burns and Anderson's (1993) theory of attentional inertia. The sustained attention scale elicited self-perceptions regarding sustained looks at and concentration on the presentation. The scale reliability (Cronbach's alpha) for the sustained attention/mental effort scale was .78.

Information Processing Quality. Perceived information processing quality was assessed using questionnaire items (semantic differential and Likert-type items) adapted from Buchheit (1996). The perceived information processing quality elicited an assessment of the perceived

interestingness, understandability, and informativeness of the presentation. The scale reliability (Cronbach's alpha) for the information processing quality scale was .90.

Comprehension. In order to facilitate objective scoring, the short essay questions were constructed and scored according to the analytic approach of essay assessment (Jacobs & Chase, 1992; Linn & Gronlund, 2000). The Photocopier Operation and Greenhouse Effect comprehension tests each required 13 correct responses. Sample comprehension questions are: "How is the electrostatic image formed on a photoconductor's drum surface?" and "How does livestock production lead to an increase in global warming?"

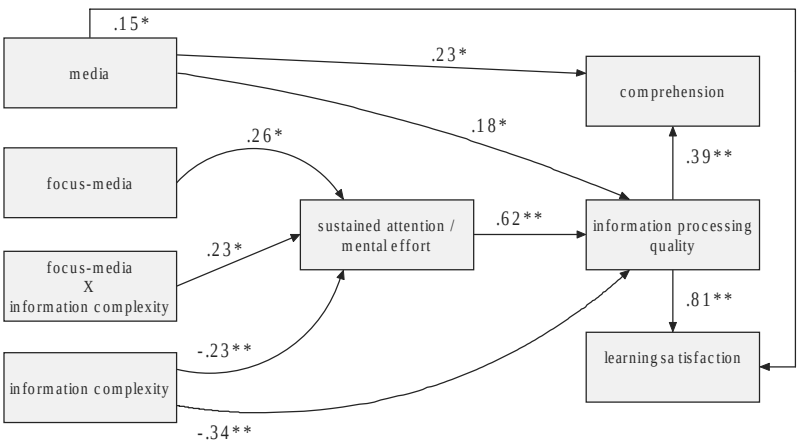
Learning Satisfaction. The learning satisfaction scale (Likert-type) was developed using Keller's (1987) ARCS Model of Motivational Design of effective learning environments. The learning satisfaction scale elicited self-perceptions regarding extent of comprehension and understanding of the information presented, sense of mastery to use the information presented, usefulness of the information presented, and desire to learn more about the information topic. The scale reliability (Cronbach's alpha) for the learning satisfaction scale was .84.

RESULTS

A series of regression analyses was chosen as the data-analytic strategy for testing the proposed causal path model depicted in Figure 1 (Pedhazur, 1997). Problem solving ability (i.e., grade point average) was considered as a covariate to comprehension but there was no significant bivariate correlation. The overall results from the series of regression analyses used to test the fit of the causal model to the data are summarized in Figure 2. Table 2 presents the means, standard deviations, original correlations, and reproduced correlations of the variables in the study.

The path coefficients (see Figure 2) were used to compute the reproduced correlation of variables (see upper half of matrix in Table 2) that are interconnected through direct and indirect causal paths (Pedhazur, 1997). The total variance explained by the causal path model was 0.927. A root mean squared residual (RMR) of 0.076 and a normed fit index (NFI) of 0.822 suggest that the model provides a good fit to the data (Gerbing & Anderson, 1993; Pedhazur, 1997).

Figure 2: Model results: Standardized regression weights.



Note: $*p < .05$ $**p < .01$

Table 2: Means, Standard Deviations, and Correlations

Variables	M	SD	1.	2.	3.	4.	5.	6.	7.	8.
1. media	♦	♦	·					(.18)	(.30)	(.29)
2. focus-media	♦	♦	.02	·			(.26)	(.16)	(.06)	(.12)
3. information complexity	♦	♦	.00	.00	·		(-.23)	(-.48)	(-.19)	(-.38)
4. focus-media x information complexity	♦	♦	.00	.16	-.06	·	(.23)	(.14)	(.06)	(.26)
5. sustained attention	18.95	5.70	-.08	.22	-.22	.18	·	(.70)	(.24)	(.51)
6. information processing quality	41.14	9.06	.13	.09	-.48**	.11	.68**	·	(.43)	(.83)
7. comprehension	3.90	2.33	.23*	.06	-.60**	.02	.24*	.43**	·	
8. learning satisfaction	16.60	4.46	.25*	.15	-.33**	.23*	.64**	.81**	.32**	·
9. problem solving ability	3.32	.35	-.22	.04	-.10	.16	.15	.04	.20	-.07

Notes $N = 78$ * $p < .05$ ** $p < .01$ ♦ media, focus-media, and information complexity are contrast coded variables
The original correlations are reported in the lower half of the matrix. The reproduced correlations of relevant variables are reported in the upper half of the matrix.

Sustained Attention and Mental Effort

As posited in Hypothesis 1, the text-only and multimedia groups (i.e., focus-media) did exhibit a positive increase in sustained attention and mental effort (standardized beta = .26, $p < .05$). Increases in information complexity lowered the motivation to maintain sustained and effortful processing of the information presented (standardized beta = -.23, $p < .01$) thereby supporting Hypothesis 2. Hypothesis 3 predicted that text-only and multimedia presentations would positively moderate the tendency for increases in information complexity to lead to withdrawal of sustained attention and mental effort as compared to text-with-graphics presentations. The results indicated that Hypothesis 3 was supported (standardized beta = .23, $p < .05$).

Information Processing Quality

The results indicated support for Hypothesis 4 and suggest that presentations that target both the verbal and visual working memory channels are perceived to be more understandable, informative, and interesting (standardized beta = .18, $p <$. 05). The results for Hypothesis 5 indicated that sustained attention and applied mental effort is positively correlated with information processing quality (standardized beta = .62, $p <$ .01). For Hypothesis 6, the results indicated that information complexity was negatively related to information processing quality (standardized beta = -.34, $p < .01$).

Comprehension and Learning Satisfaction

For Hypothesis 7, the results indicated that information processing quality was positively related to comprehension (standardized beta = .39, $p <$ .01). As suggested in Hypothesis 8, increased comprehension was associated with increased presentation modality (standardized beta = .23, $p < .05$). The results also indicated that increased information processing quality was positively related to learning satisfaction (i.e., perceived understanding and mastery of the topic). Consequently, Hypothesis 9 was supported (standardized beta = .81, $p < .01$). Finally, hypothesis 10 predicted that increases in media modality would be positively related to greater sense of mastery of use of the information acquired (i.e., learning satisfaction). The results indicated support for Hypothesis 10 (standardized beta = .15, $p < .05$).

DISCUSSION

The path coefficient between sustained attention and focus-media interaction with information

complexity (beta = .23) was only slightly lower than the path coefficient between sustained attention and focus–media alone (beta = .26). These results suggest that adverse effects on sustained attention (Burns & Anderson, 1993) typically associated with information complexity (beta = -.23) were moderated by increased media modality. Apparently, the multimedia group subjects recalled more accurate details regarding sequence of operation and dependency among photocopier components or the process of greenhouse gas generation. Further, the results suggest that multimedia did not create a split attention situation (i.e., cognitive load due to switching attention between text and visual content) typically observed in text-with-graphics presentations. This supports the major tenet of dual coding theory (Paivio, 1986); the verbal/auditory and visual memory channels actually augmented each other.

The direct effect of media on comprehension (beta = .23) supported the contention that verbal and visual working memory effects are additive and synergistic. Sustained attention exhibited a greater association with information processing quality (beta = .62) as compared to media (beta = .23). Apparently, media's role in inducing and maintaining attention was more important to information processing quality than its role in targeting both the verbal and working memory channels. Media modality was related to learning satisfaction (beta = .15) but information processing quality imposed greater relative variation (beta = .81). It is clear that sustained attention, associated with focus media, was essential for positive information processing quality needed to enhance both comprehension and learning satisfaction.

The positive relationship between focus-media (i.e., text-only and multimedia) and sustained attention and mental effort support previous findings that text-only presentations impose a required cognitive processing effort while multimedia induced voluntary attention and mental effort (Andres, 2001; Kozma, 1991). Sustained attention and mental effort was highest for the text-only group (mean = 20.25), followed by multimedia (mean = 19.27), followed by text-with-graphics (mean = 17.08). The text-only group reported higher levels of attention and mental effort, which was likely expended to encode the information (i.e., identify noun-verb linkages and to establish local and global coherence among related clauses). Here, focus and considerable mental effort are needed to translate text into a visual simulation. There is less potential to be distracted from external stimuli under such periods of high concentration. In contrast, multimedia appeared to generate higher sustained attention through the sensory appeal of its multi-modal content (i.e., text, images, narration, and animation).

The results indicate that high information complexity was second to sustained attention in its magnitude of explained variation in information processing quality (beta = -.34). The negative association of information complexity with information processing quality is evidence that there were considerable information processing errors in identifying idea units (noun-verb linkages) and establishing causal connections among idea units to create a situation model. This result supports the "cognitive resource allocation" perspective proposed by Millis et al. (1993). According to the resource allocation perspective, the presence of excessive causal connectives and high connective spans creates ambiguities in the encoding and interpretation of incoming information. This overloaded working memory and diminished information processing quality and subsequently comprehension and learning satisfaction. The results also indicate that magnitude of relationship between information complexity (beta = -23) and focus-media (beta = .26) are fairly equal but opposite in direction. This suggests that competing external stimuli or disinterest can possibly distract a subject when comprehensibility of information is difficult. In addition, high local and global connections among concepts result in minimal situation model simulation accuracy, greater sense of confusion, and greater disinterest in presentation content.

A review of relevant literature suggested that the interaction term, focus media x information complexity, should be included in the model. As stated earlier, focus-media (i.e., text-only and multimedia) has a tendency to command greater required (e.g., text parsing) or voluntary (e.g., perceptual arousal) sustained attention and mental effort. In contrast, increased complexity has the potential to lead to withdrawal of sustained attention and mental effort through incomprehensibility and disinterest in the information presented. The results for Hypothesis 3 suggest that both the encoding process (e.g., verbal vs. verbal & visual) and semantic structure of information (i.e., causal chain length and connective span) interact in the determination of effort to encode incoming information. Apparently, the semantic structure of information should drive information presentation format decisions.

IMPLICATIONS AND CONCLUSIONS

The results indicated implications for the use of computer-based instruction, meeting presentations, and computer-supported decision-making contexts. It is clear that multimedia provides a unique opportunity to examine the nature of human information processing (i.e., encoding, situation model construction, and situation model simulation) and how information processing interacts with information format and information complexity. The ability of multimedia to target both the verbal and visual working memory channels affords the opportunity to create experimental tasks that can be used to assess cognitive processing limitations. Apparently, increasing presentation modality can minimize the occurrence of excessive cognitive load when processing complex information. Multimedia-based visuals supported by text and/or narration can convey a maximum amount of information that is processed more efficiently than static images and text.

Information content should be organized so as to minimize local and global causal dependencies among concepts. Instructional and training presentations should be examined to identify what content is best suited for a specific presentation modality. Such decisions would be driven by the complexity of the topic, the extent of mental model construction and simulation required to understand the topic (i.e., abstract level), and the extent to which information processing requires greater short-term memory capacity and/or long-term memory capacity. In decision-making contexts, subjects must often analyze data, construct mental models of alternative solutions, simulate these alternative models as potential scenarios, and finally compare the outcomes of each simulation against a set of decision-making criteria. In engineering contexts (e.g., mechanical or chemical), multimedia can be used to depict physical models that represent abstract models defined by mathematical equations. It is clear that multimedia can be applied in such contexts with the aim of reducing cognitive load through the provision of graphs, imagery, and animation that summarize data physical form, and motion.

Past research on multimedia use has noted that there can be negative outcomes during multimedia use. Split-attention effect (i.e., the need to integrate information from two different sources – text and diagrams) has been observed in some multimedia studies. Split-attention was not observed in this study, and it is believed that split-attention can be avoided through careful embedding of text onto or nearby the related images. Further, some research studies on presentation mode and cognitive processing outcomes have observed mixed results regarding effectiveness of increased presentation modality. These mixed results could possibly be attributed to inconsistencies in the experimental task used (e.g., word recognition, word-image associations, short passages, minimal information complexity). In an attempt to isolate presentation/instructional media effects, presentation content across media

(i.e., text-only, overhead slides, multimedia) was structurally equivalent in idea units expressed and sequence of events. Information complexity was limited to the causal chain lengths, and local and global connections of idea units (i.e., connective span) needed to describe the *Greenhouse Effect* or *Photocopier Operation*. Readability scores were essentially equivalent.

A limitation of the current study is the use of self-reported measures for sustained attention and mental effort and information processing quality. Objective measures of sustained attention would offer greater internal validity. Future studies should attempt the use of objective measures of attention and applied mental effort. Although larger and more realistic than tasks previously utilized, this study's task can be viewed as a limitation to the study. Future studies should extend the length and complexity of the tasks in order to afford greater generalizability. Finally, the use of a laboratory study offers greater internal validity, but its use can be viewed as a limitation because field studies offer more external validity. The ability to take advantage of the power of computer-based technologies such as multimedia will depend on continued research aimed at understanding the relationship between the capabilities of the technology and information processing requirements of the task to be supported with the technology.

REFERENCES

Andres, H. P. (2001). Presentation media, information complexity, and comprehension. *Journal of Educational Technology Systems, 30(3),* 225-246.

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of medium on recall. *Journal of Verbal Learning and Behavior, 18(3),* 333-356.

Buchheit, N. A. (1996). *Multimedia: a persuasion tool for affecting decision outcome*. Unpublished Ph.D. Dissertation, Texas A&M University College of Business, College Station, TX.

Burns, J. J. and Anderson, R. A. (1993). Attentional inertia and recognition memory in adult television viewing. *Communication Research, 20(* 6), 777-799.

Cassady, J. (1998). Student and instructor perceptions of the efficacy of computer-aided lectures in undergraduate university courses. *Journal of Educational Computing Research, 19(2),* 175-189.

Christoph, R. T., Schoenfeld, G.A. and Tansky, J.W. (1998). Overcoming barriers to training utilizing technology: the influence of self-efficacy factors on multimedia-based training receptiveness. *Human Resource Development Quarterly, 9(1),* 25-38.

Cole, P. (1992). Constructivism revisited: A search for common ground. *Educational Technology, 32(2),* 27-34.

Friedman, N. P. and Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General, 129(1),* 61-83.

Gerbing, D. and Anderson, J. (1993). A Monte Carlo evaluation of goodness-of-fit indices for structural equation models. In K. A. Bollen and J. S. Long (Eds), *Testing Structural Equation Models* (pp. 40-65). Newbury Park, CA: Sage Publications.

Gordon, P. C., Hendrick, R. and Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13(5),* 425-430.

Hackman, M. Z. and Walker, K. B. (1990). Instructional communication in the televised classroom: The effects of system design and teacher immediacy on student learning and satisfaction. *Communication Education, 39(3*), 196-206.

Halford, G., Wilson, W. H. and Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral & Brain Sciences, 21(6),* 803-864.

Hegarty, M., Narayanan, N., and Freitas, P. (2002). Understanding machines from multimedia and hypermedia presentations, In J. Otero and J.A. Leon (Eds.), *The psychology of science text comprehension* (pp. 357-384). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Jacobs, L. C. and Chase, C. I. (1992). *Developing and using tests effectively: A guide for faculty.* San Francisco : Jossey-Bass Publishers.

Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99(1),* 122-149.

Keller, J.M. (1987). Strategies for stimulating the motivation to learn. *Performance and Instruction, 26(8),* 1-7.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61(2),* 179-211.

Levin, J., Anglin, G. and Carney, R. (1987). On empirically validating functions of pictures in prose. In D. Willows and H. Houghton (Eds.), *Psychology of Illustration: Vol. 2 Instructional Issues* (pp. 51-85). New York: Springer-Verlag.

Lim, K. H. and Benbasat, I. (2000). The effect of multimedia on perceived equivocality and perceived usefulness of information systems. *MIS Quarterly, 24(3),* 449-471.

Linn, R. L. and Gronlund, N. E. (2000). *Measurement and Assessment in Teaching* (8th Ed.). Upper Saddle River NJ: Prentice Hall.

Maki, R. H., Maki, W. S., Patterson, M. and Whittaker, D. (2000). Evaluation of a web-based introductory psychology course: Learning and satisfaction in on-line versus lecture courses. *Behavior Research Methods, Instruments, & Computers, 32(2),* 230-239.

Mayer, R. E and Chandler, P. (2001). When learning is just a click away: does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93(2),* 390-397.

Mayer, R. E. and Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90(2),* 312-320.

Mayer, R. E. and Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning & Instruction, 12(1),* 107-119.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29(2),* 111-123.

Millis, K. K., Golding, J. M., and Baker, G. (1995). Causal connectives increase inference generation. *Discourse Processes, 20(1),* 29-49.

Millis, K. K., Graesser, A. C. and Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology, 7(4),* 317-339.

Moreno, R. and Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94(1),* 156-163.

Mousavi, S. Y., Low, R., and Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87(2),* 319-334.

Naglieri, J. A. and Das, J. P. (1997). Intelligence revised: The planning, attention, simultaneous, successive (PASS) cognitive processing theory. In R.F. Dillon (Ed.), *Handbook on Testing* (pp. 136-163). Westport, CT: Greenwood Press.

Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford, England: Oxford University Press.

Park, O. (1998). Visual display and contextual presentations in computer-based instruction. *Educational Technology Research and Development, 46(3),* 37-50.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research* (3rd ed.). Fort Worth, TX: Harcourt Brace.

Ramalingam, V. and Wiedenbeck, S. (1998). Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. *Journal of Educational Computing Research, 19(4),* 367-381.

Rende, B., Ramsberger, G. and Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology, 16(3),* 309-321.

Rieber, L. P. (1991). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology, 83(3),* 318-328.

Song, S. H. and Keller, J. M. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational Technology Research & Development, 49(2),* 5-22.

Small, R. V. and Gluck, M. (1994). The relationship of motivational conditions to effective instructional attributes: A magnitude scaling approach. *Educational Technology, 34(10),* 33-40.

Towler, A. J. and Dipboye, R. L. (2001). Effects of trainer expressiveness, organization, and trainee goal orientation on training outcomes. *Journal of Applied Psychology, 86(4),* 664-673.

van Merrienboer, J. J., Schuurman, J. G., de Croock, M. B. and Paas, F. G. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning & Instruction, 12(1),* 11-37.

Wagenaar, W. A., Schreuder, R., and Wijlhuizen, G. J. (1987). Readability of instructional text, written for the general public. *Applied Cognitive Psychology, 1(3),* 155-167.

Washburn, D. A. and Putney, R. T. (2001). Attention and task difficulty: When is performance facilitated? *Learning & Motivation, 32(1),* 36-47.

Zwaan, R. A., Magliano, J. P., and Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(2),* 386-397.

This work was previously published in Information Resources Management Journal, Vol. 17, No. 1, edited by M. Khosrow-Pour, pp. 63-78, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.18
Interface Design, Emotions, and Multimedia Learning

Chaoyan Dong
New York University, USA

ABSTRACT

In social psychology, "what is attractive is good" means that a physically attractive person is perceived to be more favorable and capable. In industrial design, the interface is one of the three elements that influence users' experience with a product. For multimedia learning, does the interface design affect users' experience with learning environments? Does attractive interface enhance multimedia learning? Research in multimedia learning has been neglecting this issue. In this chapter, I propose that attractive interface design does indeed promote multimedia learning. This hypothesis is based on the review of the following theories and related empirical studies: (1) an interface impacts a user's experience; (2) beautiful interface induces positive emotions; (3) positive emotions broaden cognitive resources; and (4) expanded cognitive resources promote learning. The model of emotional design in multimedia learning is proposed to highlight how emotions regulate multimedia learning. Suggestions regarding designing attractive interfaces are provided.

INTRODUCTION

Multimedia learning refers to learning from multimedia design, which is the presentation of materials both in words and pictures. Multimedia design has been widely used in educational settings. Research on multimedia learning has been looking at how to design effective and efficient multimedia environments. For example, in *multimedia learning*, thus far the most comprehensive research on multimedia learning, Mayer (2001) summarizes seven multimedia learning principles, that is, spatial contiguity principle, multimedia principle, temporal contiguity principle, coherence principle, modality principle, redundancy principle, and individual difference principle. All of these principles are about the design of text, audio, and video, each of which is assumed to be a multimedia design element that determines the results of multimedia learning. Unfortunately, the assumption is only partially true when the design is always for one group of learners. In reality, the idea of "one size fits all" probably never works. It is critical to consider the roles of both multimedia

designers and learners when talking about the quality of multimedia learning.

Multimedia designers determine interface design in addition to texts, audio, and video. Interface design refers to designing the interaction between a human and a machine (Raskin, 2000). The interface design induces certain emotions from users while they interact with the design. In other words, interface design is the visible surface that users experience while interacting with a design, while emotions are the underlying, invisible media between the users and the design. Research on emotions indicates that emotions play as important a role as cognition does in learning. It is widely agreed that positive emotions enhance cognitive activities, although the cognitive activities do not necessarily entail learning or multimedia learning. Therefore, when talking about the quality of multimedia learning, we must address the issue of how the interface design affects multimedia learning and should consider the emotions induced from experiencing the multimedia design. The discussion in this chapter helps to identify interface design and emotions as influences in multimedia design that is not subsumed by the influences on efficiency and effectiveness that have traditionally been researched by multimedia learning theorists.

The following section explains the theoretical framework of how interface design affects users' experience as well as their emotional states, especially how positive emotions influence cognition, and how changes in cognition regulate multimedia learning. Based on the theoretical framework, the emotional design model in multimedia learning is proposed, which is the main focus of the chapter. Since positive emotions facilitate cognitive activities, as suggested by the theoretical framework, design features that intend to induce positive emotions are discussed. Future trends in research of emotional design in multimedia learning are also discussed.

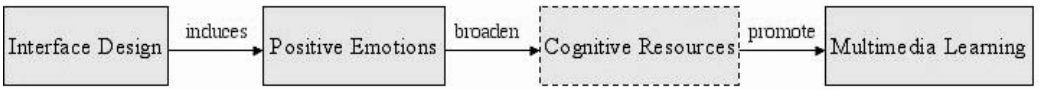
BACKGROUND

Interface design is the first thing users experience when interacting with a multimedia design. Emotions are induced before initiation of cognitive activities to process in users' brains. In other words, interacting with the interface design induces emotions and also activates cognitive activities from users. Emotional change is a rapid activity, preceding the cognitive activities. Norman's (2004) theory of emotional design proposes a theoretical framework to explain how to interact with an interface design affects users' emotions, and also suggests that attractive designs induce positive emotions from users. Fredrickson's (1998) positive emotion theory elucidates how positive emotions facilitate cognitive activities. The goal for multimedia learning research is to make the learning effective and efficient. We should ask, Do positive emotions promote multimedia learning? Mayer's cognitive theory of multimedia learning explains the general process of multimedia learning. One of Mayer's assumptions is that the working memory has a limited capacity, but he does not consider the possibility that positive emotions broaden cognitive resources. Does it mean that positive emotions promote multimedia learning by expanding the working memory capacity? The discussion is illustrated in Figure 1 by connecting the four theories, which are combined to form the conceptual framework for the chapter. The details of each theory and the connections between these theories are explained in the following section.

How Does the Interface Design Affect Users' Experience?

Norman (2004) proposes a framework to describe what happens when users interact with a design. According to Norman, a product presents three aspects to its users: its attractiveness, its usability, and the reflective images. The three features induce different emotions from users. Attractiveness

Figure 1. The conceptual framework



is a result of the visceral level activity in the brain. Visceral activity is a "rapid, reactive response to appearance" (Norman, p. 21), and induces taste-based emotions from users (e.g., attractiveness or unattractiveness). Norman suggests that the visceral design be developed to match the visceral activity in the brain. Visceral design is the design of a product's appearance and feel. Usability is determined by the brain level that processes everyday behaviors. Usability induces goal-based emotions (e.g., satisfaction, distress, optimistic expectation, worry, relief, frustration, and disappointment). Behavioral design concerns how easy the design is to use. Experimental studies (Kurosu & Kashimura, 1995; Tractinsky, 1997; Tractinsky, Katz, & Ikar, 2000) indicate that interface aesthetics could really be important and positively impact the perceived usability of a design. The reflective image is the result of reflective activity in the brain. Reflectivity induces standard-based emotions such as admiration, gratitude, pride, anger, and resentment.

Ortony's (2003) theory is similar to Norman's in the sense that both are based on a framework that explains users' interaction with a design. The key difference is that Ortony puts more emphasis on the factors that induce users' emotions while interacting with a product. He points out that the designers' motivation in designing a product also determines users' emotions in addition to the users' approach. He categorizes designers' motivation as indifference, prevention, and promotion. When designers are indifferent, it is possible that no emotions will be induced from users, except perhaps by accident; Ortony does not include emotions induced by accident in his theory. When designers aim for prevention, the product will probably not induce emotions from users, as these are the result of design. It is when designers aim for promotion that the product is most likely to induce emotions from users, in this case by design.

In summary, both Norman (2004) and Ortony (2003) suggest that the interface design induces emotions from users. Experimental studies are available to support the argument (Klein, Moon, & Picard, 2002; Kurosu et al., 1995; Lester et al., 1997; Tractinsky, 1997; Tractinsky, Katz & Ikar, 2000). For example, Desmet (2002) defines 41 design-related emotions and the factors that elicit these emotions. Attractiveness, one of the design-related emotions, is induced by one specific feature of the interface or by the overall appearance. The underlying assumption is that attractive interface induces positive emotions, which is consolidated in his later research.

What is Attractive is Good?

Attractiveness in daily language corresponds to aesthetics in philosophy, so a design that represents fun, cuteness, and includes color may be considered aesthetic if the user likes it. Since experiencing a multimedia design induces certain emotions from users, does attractive interface design induce positive emotions from users as suggested by Norman's (2004) theory? LeDoux (1982) proposes a dual-processing theory for visual images. After a visual design registers in the brain, it is processed either in the format of unconscious emotions or in a detailed perceptual analysis. Both processes happen in two different regions of the brain, that is, the thalamo-amygdala pathway and the cortical pathway. It means that a

visual design induces both emotional responses and cognitive activities in the brain. On the other hand, research on visual design indicates that users form the emotional state of "like" while considering a design as attractive. "Like" is one category of positive emotions. This phenomenon is supported by the implicit personality theory in social psychology (Schneider, 1973), which explains how human beings perceive other people. The assumption of the theory is that human beings are "naïve scientists" actively perceiving their surroundings. The process of perceiving other people includes describing the attributes of personality that an individual believes others to possess, predicting the relations between these attributes, and explaining why one person behaves in certain ways. The prediction of the theory is that physically attractive people are believed to be more capable (Ashmore, 1981; Dion, Berscheid, & Walster, 1972; Eagly, Ashmore, Makhijani, & Longo, 1991). It conforms to Norman's (2004) proposition that attractive appearances induce positive emotions. In summary, the dual-processing theory for visual images, the research on visual design and the implicit personality theory, collectively indicate that attractive interface design induces positive emotions in users, which is supported by experimental studies. For example, Schenkman and Jonsson (2000) discover that people prefer aesthetically pleasing Web sites, which induce positive emotions from users. Yamamoto and Lambert (1994) investigate how a product's aesthetics impacts users' evaluation of industrial products and find that product appearance has a moderate impact on customers' preference. Jordan's (1998) study indicates that the aesthetically pleasing products induce positive emotions; participants will use more aesthetically pleasing products than those unattractive products. Lavie and Tractinsky (2004) suggest that "the visual aesthetics of computer interfaces is a strong determinant of users' satisfaction and pleasure" (p. 269). Van der Heijden's (2003) study concludes that the perceived visual attractiveness of the interface positively impacts users' attitudes and intention toward a design, which positively impacts actual usage. Demirbilek and Sener (2001) review literature on product design and emotions and conclude that certain design elements (i.e., fun, cuteness, familiarity, metonymy, and color) induce positive emotions from users. These findings are confirmed in a later experimental study by Demirbilek and Sener (2003).

Positive Emotions

Attractive design induces positive emotions from users, but how are positive emotions defined? And how do positive emotions impact cognition? Positive emotions are a category of emotions, sharing features identified by the following theories on emotions: emotions refer to mental states (Cornelius, 1996). The cognitive perspective of emotions focuses on the role that thought plays in the process of emotions (Arnold, 1960; Frijda, 1986; Lazarus, 1991; Oatley, 1992; Oatley & Johnson-Laird, 1987; Zajonc, 1980). Smith and Lazarus (1993) propose a cognitive-motivational relational theory, which claims that emotions are preceded by appraisal triggered by specific environments and related to an individual's experience. Positive emotions are not simply the opposite of negative emotions (e.g., that happiness and sadness are controlled by independent neural pathways) (George et al., 1995). A growing body of empirical evidence shows that positive and negative emotions have qualitatively different information-processing models (Gray, 2001; Isen, 1999; Kuhl, 1983, 2000). Therefore, positive emotions and negative emotions play different roles in cognitive processes, with positive emotions playing a particularly important role (Diener & Larsen, 1993; Myers & Diener, 1995). Fredrickson's (1998) broaden-and-build theory of positive emotions provides a framework for understanding how positive emotions impact cognitive processes. According to Fredrickson, positive emotions broaden the thought-action repertoire. Specifically, positive emotions broaden

the scope of attention, cognition, and action, as well as building physical, intellectual, and social resources. The outcome of the broadened thought-action repertoires is an increase in physical, intellectual, and social resources. The increase in these resources as a result of experiencing positive emotions is stable, outlasting the transient emotional states. Fredrickson defines the contexts that induce positive emotions as safety and satiation. Fredrickson's theory is supported by Isen and her colleagues' experimental studies on positive emotions. These demonstrate positive emotions linked to an increase in brain dopamine levels, so the authors conclude that "the elevated dopamine levels influence performance on a variety of cognitive tasks" (Ashby, Isen, & Turken, 1999, p. 544).

A large amount of research has shown convincingly that positive emotions systematically influence performance on many cognitive tasks, which supports Fredrickson's statement that positive emotions promote cognitive activities. For example, positive emotions improve creative problem solving (Estrada, Isen, & Young, 1994; Greene & Noice, 1988, Isen, Daubman, & Nowicki, 1987), enhance recall of study material (Isen, Shalker, Clark, & Karp, 1978; Lee & Sternthal, 1999), and systematically change strategies used in decision-making tasks (Carnevale & Isen, 1986; Estrada, Isen, & Young, 1997). Bolt, Cogschke, and Kuhl's (2003) study indicates that positive emotions improve participants' judgments. Fredrickson and Branigan's (2005) study validates the broaden-and-build theory that positive emotions broaden the scope of attention and thought-action repertoires, whereas negative emotions narrowed thought-action repertoires. Gasper (2004) tests the level of focus hypothesis proposed by Fredrickson (1998) that a happy person is more likely to process stimuli as a whole. The results suggest that happier individuals process information more globally than do those in negative moods.

In summary, positive emotions generally promote cognitive activities, and interacting with attractive interface design induces positive emotions from users. How do in particular positive emotions promote learning from multimedia? Before answering the questions, let us first review Mayer's (2000) cognitive theory of multimedia learning, which explains how multimedia learning happens.

Mayer's Cognitive Theory of Multimedia Learning

Three assumptions are employed for the theory, that is, the dual-channel assumption, the limited-capacity assumption, and the active-processing assumption. The dual-channel assumption states that the sensory modes for information input include two channels: ears processing verbal information and eyes for pictorial information. The limited-capacity assumption is closely related to the model of working memory by Baddeley (1986, 1992, 1999) and the cognitive load theory by Chandler and Sweller (1991; Sweller, 1999). The limited-capacity assumption states that the amount of information processed by each channel at one time is limited. According to Miller (1956), the average amount of information processed at each channel at one time is five to nine chunks. The active-processing assumption proposes how the human brain processes information, that is, selecting information, organizing the incoming information, and integrating the information with other knowledge stored in the long-term memory. Based on the three assumptions, the cognitive theory of multimedia learning proposes that multimedia information is presented in two formats, words, and pictures. Verbal information enters working memory through the ears, while visual information enters working memory through the eyes. In working memory, verbal information interacts with visual information, and the information is organized into either verbal models or pictorial models. The verbal models and the pictorial models are integrated with individual's prior knowledge of the specific topic. In long-term

memory, the integrated information from verbal model, pictorial model, and prior knowledge is formed as schemata and stored in long-term memory. The cognitive theory of multimedia learning is supported by numerous experimental studies by Mayer and his colleagues (Harp & Mayer, 1997; Mayer, 1996, 1997; Mayer & Anderson, 1991; Mayer & Gallini, 1990; Mayer & Sim, 1994; Moreno & Mayer, 1999).

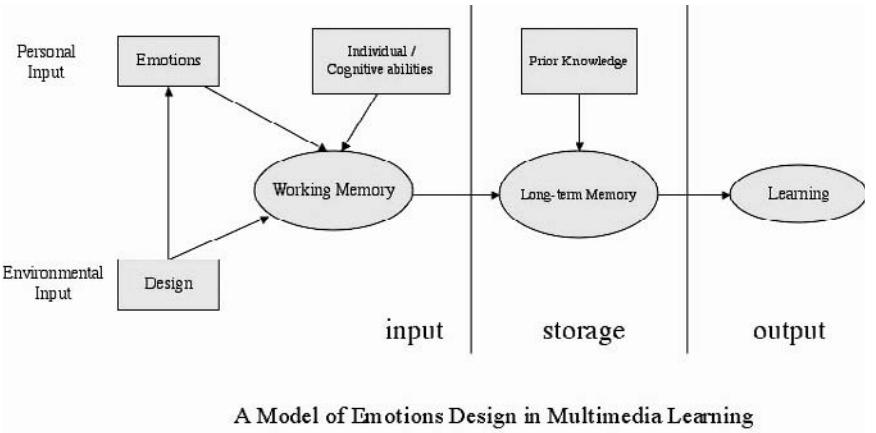
According to Mayer (2001), the capacity of the working memory and the capacity of both auditory and visual channels are limited. Fredrickson (1998) explains that positive emotions broaden the cognitive resources. However, she does not explain whether cognitive resources are related to the capacity of working memory. Further research is needed to verify whether positive emotions expand working memory capacities and whether positive emotions promote multimedia learning like that positive emotions do with other cognitive activities. One possible interpretation is that positive emotions increase the cognitive capacity and the working memory, which results in improved learning. The second possible interpretation is that the increased cognitive resources increase the amount of information processed in the working memory, which finally promotes learning. In summary, the four theoretical perspectives indicate that an attractive interface should promote multimedia learning.

MODEL OF EMOTIONAL DESIGN IN MULTIMEDIA LEARNING

The four theories presented previously provide insight into the cognitive mechanisms underlying how attractive design affects multimedia learning through positive emotions. Based on the preceding discussion, we introduce the model of emotional design in multimedia learning, a framework for integrating emotions into a multimedia learning environment. Later the implications of the model for multimedia design are discussed.

The model of emotional design in multimedia learning is intended to highlight the impact of emotions on multimedia leaning (see Figure 2). The model is built on three previous theories (i.e., the information-processing theory in cognitive psychology (Miller, 1956), the cognitive theory of emotions (Arnold, 1960), and the cognitive theory of multimedia learning (Mayer, 2001)). The information-processing theory proposes that the structure of human cognition includes both information input and output systems. Information input includes learners' input and environmental

Figure 2. A model of emotions design in multimedia learning



input. The environmental input refers to the multimedia design (Norman, 2004). Personal input includes the emotions that are induced from multimedia design, individual/cognitive abilities, and prior knowledge. Arnold claims that the sequence of emotional processes is as follow: perception → appraisal → emotions. Appraisal refers to the process of judging how important an event is to a person. Emotions are induced when one person encounters an event and judges how important the event is to him or her, but if aesthetic enough, need not be important. In multimedia learning situations, emotions are induced when users interact with a multimedia design. As discussed previously, Mayer's cognitive theory of multimedia learning explains how multimedia learning occurs through both verbal and audio channels, in working memory and in long-term memory. However, the amount of information processed by each channel at one time is limited. The capacity of working memory is also limited. Do positive emotions affect the capacity of the two information-processing channels and the capacity of the working memory? Further research is needed to answer the question.

The model suggests a set of prescriptive propositions that are obtained from deductively and inductively established relationships between related concepts, theories, and empirical research in the field of emotions, aesthetics, learning, and multimedia design. Prescriptive propositions point out what should be done by the multimedia designers during design or what kind of strategies should be applied in order to elicit the desired emotional outcomes from the users.

Suggestions for Attractive Multimedia Design

There clearly exist individual and cultural differences regarding perceived attractiveness. However, common psychological mechanisms shared by all human beings underlie aesthetics that can be incorporated into multimedia interfaces. A review of research on aesthetics and graphic design indicates the design elements related to multimedia aesthetic, including color, graphics, text, audio, and video.

1. **Color:** Colors in graphics serve informational, compositional, and expressive functions, which black-and-white designs do not possess (Zetll, 2005). Color energy refers to users' aesthetic responses to a color. The energy of a color is determined by "(1) the hue, saturation, and brightness of a color; (2) the size of the colored area; and (3) the relative contrast between foreground and background colors" (Zetll, p. 67). Saturation influences color energy most. High saturation means high energy, and vice versa. High-energy warm colors generally induce a happier mood on users than do low-energy cool colors. High brightness colors have higher energy than low brightness colors. The color combination of small areas of high-energy colors against large background areas of low-energy colors is perceived as pleasant. The most pleasant background hues are blue, blue-green, green, red-purple, purple, and purple-blue (Valdez & Mehrebian, 1994).
2. **Graphic:** Gestalt theory claims that different elements, when combined as a whole, reveal more information than elements viewed in isolation (Wertheimer, 1923). The practical implication of the Gestalt theory for graphic design is emphasis on relationships between different well-designed elements. The purpose of the design is to reflect abstract scientific concepts and relationships thereby helping users to create accurate mental models of them. The ratio between an element and its context should reflect the actual ratio so that the graphics portray relationships precisely. For example, the Earth and Sun's relative sizes and separating distances should be accurately reflected in illustrated

objects and their parts (Holliday, 2001). The details revealed by a graphic affects users' interpretations of the graphic. Too much information distracts users from essential information because their eyes might not know where to go (Wittich & Schuller, 1967). In a pace-controlled learning environment, graphics with relatively small amounts of information (e.g., simple line drawings) tend to be more effective (Dwyer, 1972). Most computer displays follow a 3:4 aspect ratio so the screen area of a design should also follow an approximate 3:4 ratio. According to Zettl (2005), a 3:4 ratio ensures that the difference between height and width is not obvious and users will not emphasize one dimension over the other. Graphic conventions familiar to target audiences should be used.

3. **Text:** Borchardt (1999) creates a vision scheme for the design of texts. The scheme includes size, locale, proportionality, color, and contrast of texts. He points out that fonts should be legible and in proportion to the graphics. Zettl (2005) emphasizes the importance of continuity, which means that the text should maintain its colors and size throughout the instruction. Contrast between the colors of texts and their background should remain the same as the contrast between the graphics and background.
4. **Animation:** According to Holliday (2001), it is necessary to "highlight, reintegrate, reinforce, and rehearse" some parts of the graphic design to effectively explain a scientific phenomenon (p. 8). For example, slow motion is used to highlight certain parts of an animation. To achieve slow motion, frame density is increased, that is, the motion is divided into more frames during the actual filming). In animation design, slow motion animation also runs through more frames per second than normal.
5. **Audio:** According to Borchardt (1999), factors that influence the quality of sounds include volume, pitch, timbre, attack and decay, rhythm, duration, velocity, acceleration, iteration, periodicity, familiarity, and predictability. Borchardt integrates these factors and creates an audio scheme to evaluate the quality of audio. He puts each of the factors on a spectrum. The left-hand side of the spectrum indicates low and the right-hand side for high. In general, when these factors are in the middle of a spectrum, the audio is most pleasant. Zettl (2005) proposes that sound aesthetics are determined by perspective, continuity, and picture-sound combination. Perspective means to match louder sounds with close-up pictures and far-away sounds with long shots. Sound continuity means that the sound retains its volume and quality. Another factor is picture-sound combination. I suggest using a homophonic video-audio structure, because multimedia learning is more effective when corresponding audio and video information is presented simultaneously (Mayer, 2001).

FUTURE TRENDS

Although interface design has been shown to impact multimedia learning by moderating users' emotions, there are many unanswered questions about how to define attractive interfaces and how to design attractive interface that account for individual differences in aesthetic perception. Although common aesthetic criteria exist, individual differences are still important. Philosophers have researched aesthetics for more than two centuries, so extensive literature review on aesthetics should be conducted to enlighten the research of attractive interface design. In industrial design, designers also have been trying to create attrac-

tive design for a long time and have proposed practical suggestions to make designs more attractive. Future research is needed to combine the results from aesthetics and industrial design and, in particular, to promote attractive design in learning environments.

Research on emotions, including the benefits of positive emotions, has also been conducted for more than one century. However, the mechanism underlying how positive emotions stimulate cognition is not yet clear. This explains the lack of consensus among psychologists regarding the relationship between positive emotions and cognition. Fredrickson's (1998) theory is theoretically sound, but not sufficiently supported by experiments, so her theory of positive emotions requires further exploration. More experimental studies specifically addressing the affect of positive emotions on learning particularly in the fast global Internet world are needed.

Educators have been training students to help them understand their emotions and to control their emotions, and researchers have evaluated their success. However, educators have generally neglected to design learning materials and learning environments that promote positive emotions specifically to enhance learning. In multimedia learning environments specifically, none of these issues have been adequately addressed, especially with respect to immersive, innovative technologies. Future research should also consider other aspects of learning materials and environments that impact students' emotions.

Research on how people learn indicates that the following elements impact learning, that is, teachers, instructional strategies, learners, learning materials, learning environments, and learning strategies. As discussed in this chapter, interface design is a part of the learning environment that impacts users' emotions, and emotions are believed to greatly impact learning. So what aspects in the previously mentioned elements will most significantly enhance learning? This more general and essential question should be addressed in the future research.

CONCLUSION

In his article "*Multimedia Learning: Are We Asking the Right Questions*?" Mayer (1997) noted, "At this time, the technology for multimedia education is developing at a faster pace than a corresponding science of how people learn in multimedia environments" (p. 4). Now, 10 years later this is even more true! It is essential for researchers to understand how people learn in multimedia environments. From the instructional designers' perspective, it is also important to understand how different designs impact users' emotions, and how the induced positive emotions promote learning. Multimedia development should be informed by theories tested by experimental research so that multimedia designs that predictably induce positive emotions and thereby promote multimedia learning can be developed.

REFERENCES

Arnold, M. B. (1960). *Emotion and personality: Vol. 1. Psychological aspects.* New York: Columbia University Press.

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*(3), 529-550.

Ashmore, R. D. (1981). Sex stereotypes and implicit personality theory. In D. L. Hamilton (Eds.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 37-81). Hillsdale, NJ: Erlbaum.

Baddeley, A. D. (1999). *Human memory.* Boston: Allyn & Bacon.

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556-559.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Oxford University Press.

Bolte, A., Coschke, T., & Kuhl, J. (2003). Emotion and intuition: Effects of positive and negative mood on implicit judgments of semantic coherence. *Psychological Science, 14*(5), 416-421.

Borchardt, F. L. (1999). Towards an aesthetics of multimedia. *Computer Assisted Language Learning, 12*(1), 3-28.

Brink, T., Gergle, D., & Wood, S. D. (2002). *Usability for the Web: Designing Web sites that work.* San Francisco: Morgan Kaufmann Publishers.

Budd, M. (1995). *Values of art.* London: Penguin Books.

Carnevale, P. J. D., & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behavior and Human Decision Processes*, 37, 1-13.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293-332.

Cornelius, R. R. (1996). *The science of emotion: Research and tradition in the psychology of emotion.* Upper Saddle River, NJ: Prentice Hall.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428-434.

Demirbilek, O., & Sener, B. (2003). Product design, semantics, and emotional response. *Ergonomics, 46*(13/14), 1346-1360.

Demirbilek, O., & Sener, B. (2001). A design language for products: Designing for happiness. In M. G. Helander, H. M. Khalid, & M. P. Tham (Eds.), *Proceedings of the International Conference on Affective Human Factors Design* (pp. 19-24). London: ASEAN Academic Press.

Desmet, P. (2002). *Designing emotions.*

Diener, E., & Larsen, R. J. (1993). The experience of emotional well-being. In M. Lewis, & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 405-415). New York: Guilford.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285-290.

Dwyer, F. M. (1972). *A guide for improving visualized instruction.* University Park, PA: State College, Pennsylvania State University, Learning Services Division.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but …: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109-128.

Estrada, C. A., Isen, A. M., & Young, M. J. (1994). Positive affect influences creative problem solving, and reported source of practice satisfaction in physicians. *Motivation and Emotion*, 18, 285-299.

Estrada, C. A., Isen, A. M., & Young, M. J. (1997). Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organizational and Human Decision Processes, 72*, 117-135.

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology, 2*(3), 300-319.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and though-action repertoires. *Cognition & Emotion*, *19*(3), 313-332.

Fredrickson, B. L., Mancuso, R. A., Branigan, C., & Tugade, M. (2000). The undoing effect of

positive emotions. *Motivation and Emotion, 24*, 237-258.

Frijda, N. H. (1986). *The emotions.* Cambridge: Cambridge University Press.

Gasper, K. (2004). Do you see what I see? Affect and visual information processing. *Cognition and Emotion*. *18*(3), 405-421. Retrieved on July29, 2004, from http://titania.ingentaselect.com/vl=214834/cl=112/nw=1/rpsv/cw/psych/02699931/v18n3/s6/p405

George, M. S., Ketter, T. A., Parekh, P. I., Horwitz, B., Herscovitch, P., & Post, R. M. (1995). Brain activity during transient sadness and happiness in healthy women. *American Journal of Psychiatry, 152*, 341-351.

Gray, J. R. (2001). Emotional modulation of cognitive control: Approach-withdrawal states double-dissociate spatial from verbal two-back task performance. *Journal of Experimental Psychology: General, 130*, 436-452.

Greene, T. R., & Noice, H. (1988). Influence of positive affect upon creative thinking and problem solving in children. *Psychological Reports, 63*, 895-898.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: on the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92-102.

Hofmeester, G. H., Kemp, J. A. M., & Blankendaal, A. C. M. (1996). Sensuality in product design: A structured approach. *Proceedings for CHI' 96 Conference*. Retrieved on July 6, 2004, from: http://www.sigchi.org/chi96/proceedings/desbrief/Hofmeester/ghh_txt.htm

Holliday, W. G. (2001). Textbook illustrations: Fact of filler? *The Science Teacher, 57*(9), 27-29.

Isen, A. M. (1999). Positive affect. In T. Dalgleish, & M. Power (Eds.), *The handbook of cognition and emotion* (pp. 521-539). New York: Wiley.

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*, 1122-1131.

Isen, A. M., Shalker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology, 36*, 1-12.

Jordan, P. W. (1997). Human factors for pleasure in product use. *Applied Ergonomics, 29*(1), 25-33.

Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers, 14*, 119-140.

Kuhl, J. (1983). Emotion, Kognition und Motivation: II. Die funktionale Bedeutung der Emotionen fur das problemlosende Denken und fur das konkrete Handeln [Emotion, cognition, and motivation: II. The functional role of emotions in problem-solving and action control]. *Sprache & Kognition, 4*, 228-253.

Kuhl, J. (2000). A functional-design approach to motivation and self-regulation: The dynamics of personality systems interactions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 111-169). San Diego, CA: Academic Press.

Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. inherent usability. *Proceedings for CHI '95 Conference Companion* (pp. 292-293).

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of Web sites. *International Journal of Human-Computer Studies, 60*, 269-298.

Lazarus, R. S. (1991b). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist, 46*, 819-834.

LeDoux, J. (1982). Thoughts on the relations between emotions and cognition. *American Psychologist, 37,* 1019-1024.

Lee, A., & Sternthal, B. (1999). The effects of positive mood on memory. *Journal of Consumer Research, 26,* 115.

Lester, J. C., Converse, S. A., Kahler, S. H., Barlow, S. T., & Stone, B. A. (1997). The persona effect: Affective impact of animated pedagogical agents. *Electronic Proceedings of CHI' 97, USA.* Retrieved on July27, 2004, from http://www.acm.org/sigchi/chi97/proceedings/paper/jl.htm

Levinson, J. (1996). *The pleasures of aesthetics.* Ithaca, NY: Cornell University Press.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32,* 1-19.

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review, 8,* 357-371.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83,* 484-490.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*(4), 806-813.

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*(2), 389-395.

Mayer, R. E., & Gallini, J. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82,* 715-726.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90,* 312-320.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 84,* 389-401.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review,* 63, 81-97.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

Moreno, R., Mayer, R. E., Spires, H., & Lester, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177-214.

Myers, D. G., & Diener, E. (1995). Who is happy? *Psychological Science,* 6, 10-19.

Norman, D. (2004). *Emotional Design*. Basic Books.

Oatley, K. (1992). *Best laid schemes: The psychology of emotions.* Cambridge: Cambridge University Press.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion, 1,* 29-50.

Ortony, A. (2003). Emotion by accident, emotion by design. *Paper presented at Interaction design institute, Ivrea*.

Raskin, J. (2000). *The human interface design: New directions for designing interactive systems.* Reading, MA: Addison Wesley.

Schenkman, B. N., & Jonsson, F. U. (2000). *Behavior & Information Technology, 19*(5), 367-377.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79,* 294-309.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion, 7,* 233-269.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295-312.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

Tractinsky, N. (1997). Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. *Electronic Proceedings for CHI'97: USA.* Retrieved on July 6, 2004, from http://www.acm.org/sigchi/chi97/proceedings/paper/nt.htm

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*(2), 127-145.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General, 123*(4), 394-409.

van der Heijden, H. (2003). Factors influencing the usage of Web sites: The case of a generic portal in the Netherlands. *Information & Management, 40,* 541-549.

Walton, K. (1993). How marvelous! Toward a theory of aesthetic value. *Journal of Aesthetics and Art Criticism, 51,* 499-510.

Wertheimer, M. (1923). Laws of organization in perceptual forms. First published as Untersuchungen zur Lehre von der Gestalt II. In *Psycologische Forschung,* 4, 301-350. Translation published in Ellis, W. (1938). A source book of Gestalt psychology. (pp. 71-88). London: Routledge & Kegan Paul. Retrieved on January 10, 2006, from: http://psy.ed.asu.edu/~classics/Wertheimer/Forms/forms.htm

Yamamoto, M., & Lambert, D. R. (1994). The impact of product aesthetics on the evaluation of industrial products. *Journal of Product Innovation Management, 11,* 309-324.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151-175.

Zetll, H. (2005). *Sight, sound, motion: Applied aesthetics* (4th ed.) Belmont, CA: Thompson/Wadsworth.

KEY TERMS

Aesthetics: Both the study of beauty and the properties of a system that appeal to the senses, as opposed to the content, structures, and utility of the system itself (Budd, 1995).

Emotions: Refer to mental states (Cornelius, 1996). The cognitive perspective of emotions focuses on the role that thought plays in the process of emotions (Arnold, 1960).

Multimedia Design: The presentation of materials both in words and pictures (Mayer, 2001).

Multimedia Learning: Refers to learning from multimedia design, that is, words and pictures (Mayer, 2001).

Interface Design: Refers to how the design presents information to users so that users can process information as required to complete a task (Raskin, 2000).

Positive Emotions: A category of emotions, sharing features identified by the theories on emotions. Positive emotions promote cognitive activities.

Usability: "The degree to which people (users) can perform a set of required tasks" (Brink, Gergle, & Wood, 2002).

This work was previously published in Handbook of Research on Instructional Systems and Technology, edited by T. T. Kidd and H. Song, pp. 79-91, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.19
Incorporating and Understanding the User Perspective

Stephen R. Gulliver
Brunel University, UK

ABSTRACT

This chapter introduces a selection of studies relating to each of the multimedia senses — olfactory (smell), tactile/haptic (touch), visual (sight), and auditory (sound) — and how such studies impact user perception and ultimately user definition of multimedia quality. A model of distributed multimedia is proposed, to allow a more structured analysis of the current literature concerning video and audio information. This model segregates studies implementing quality variation and/or assessment into three discrete information abstractions (the network, media, and content levels) and from two perspectives (the technical and user perspectives). It is the objective of the author that, by placing current research in context of a quality structure, the need for fuller incorporation of the user perspective in multimedia quality assessment will be highlighted.

INTRODUCTION

Multimedia quality is a multi-faceted concept that means different things to different people (Watson & Sasse, 1997). Multimedia quality definition involves the integration of quality parameters at different levels of abstraction and from different perspectives. Indeed, the perception of multimedia quality may be affected by numerous factors, for example, delay or loss of a frame, audio clarity, lip synchronisation during speech, video content, display size, resolution, brightness, contrast, sharpness, colourfulness, as well as naturalness of video and audio content, just to name a few (Ahumada & Null, 1993; Apteker, Fisher, Kisimov, & Neishlos, 1995; Klein, 1993; Martens & Kayargadde, 1996; Roufs, 1992). Moreover, as multimedia applications reflect the symbiotic *infotainment* duality of multimedia, that is, the ability to transfer information to the

user while also providing the user with a level of subjective satisfaction, incorporating the user perspective in a multimedia quality definition is further complicated since a comprehensive quality definition should reflect both how a multimedia presentation is understood by the user, yet also examine the user's level of satisfaction. Interestingly, all previous studies fail to either measure the infotainment duality of distributed multimedia quality or comprehensively incorporate and understanding the user-perspective.

Inclusion of the user-perspective is of paramount importance to the continued uptake and proliferation of multimedia applications since users will not use and pay for applications if they are perceived to be of low quality. In this chapter, the author aims to introduce the reader to work relating to each of the multimedia senses and how such studies impact user perception and definition of multimedia quality. The author proposes a model in which quality is looked at from three distinct levels: the *network-*, the *media-* and the *content-levels*; and from two views: the *technical-* and the *user-perspective.* This model is used to help structure, specifically current sight and sound literature, in order to help outline the diverse approaches used when varying and assessing multimedia quality, and ultimately to emphasize the need for fuller incorporation of the user perspective in multimedia quality assessment.

PERCEPTUAL STUDIES AND IMPLICATIONS

In this section we aim to introduce the reader to the studies relating to the four multimedia senses that lie at the core of the human perceptual/sensory experience.

Olfactory

Research in the field of olfaction is limited, as there is no consistent method of testing user capability of smell. The first smell-based multimedia environment (sensorama) was developed by Heilig (1962, 1992), which simulated a motorcycle ride through New York and included colour 3D visual stimuli, stereo sound, aroma, and tactile impacts (wind from fans, and a seat that vibrated).

A major area of olfactory research has been to explore whether scent can be recorded and therefore replayed to aid olfactory perceptual displays (Davide, Holmberg, & Lundstrom, 2001; Ryans, 2001). Cater (1992, 1994) successfully developed a wearable olfactory display system for a fire fighters training simulation with a Virtual Reality (VR) oriented olfactory interface controlled according to the users location and posture. In addition, researchers have used olfaction to investigate the effects of smell on a participant's sense of presence in a virtual environment and on their memory of landmarks. Dinh, Walker, Bong, Kobayashi, and Hodges (2001) showed that the addition of tactile, olfactory, and/or auditory cues within a virtual environment increased the user's sense of presence and memory of the environment.

Tactile/Haptics

Current research in the field of haptics focuses mainly on either sensory substitution for the disabled (tactile pin arrays to convey visual information, vibro-tactile displays for auditory information) or use of tactile displays for teleoperation (the remote control of robot manipulators) and virtual environments. Skin sensation is essential, especially when participating in any spatial manipulation and exploration tasks (Howe, Peine, Kontarinis, & Son, 1995). Accordingly, a number of *tactile display* devices have been developed that simulate sensations of contact. While "tactile display" describes any apparatus that provides haptic feedback, tactile displays can be subdivided into the follow groups:

- **Vibration** sensations can be used to relay information about phenomena, such as

surface texture, slip, impact, and puncture (Howe et al. 1995). Vibration is experienced as a general, non-localised experience, and can therefore be simulated by a single vibration point for each finger or region of skin, with an oscillating frequency range between 3 and 300 Hz (Kontarinis & Howe, 1995; Minsky & Lederman, 1996).

- **Small-scale shape or pressure** distribution information is more difficult to convey than that of vibration. The most commonly-used approach is to implement an array of closely aligned pins that can be individually raised and lowered against the finger tip to approximate the desired shape. To match human finger movement, an adjustment frequency of 0 to 36 Hz is required, and to match human perceptual resolution, pin spacing should be less than a few millimetres (Cohn, Lam, & Fearing, 1992; Hasser & Weisenberger, 1993; Howe et al., 1995).
- **Thermal displays** are a relatively new addition to the field of haptic research. Human fingertips are commonly warmer than the "room temperature". Therefore, thermal perception of objects in the environment is based on a combination of thermal conductivity, thermal capacity, and temperature. Using this information allows humans to infer the material composition of surfaces as well as temperature difference. A few thermal display devices have been developed in recent years that are based on Peltier thermoelectric coolers, solid-state devices that act as a heat pump, depending on direction of current (Caldwell & Gosney, 1993; Ino, Shimizu, Odagawa, Sato, Takahashi, Izumi, & Ifukube, 2003).

Many other tactile display modalities have been demonstrated, including *electrorheological devices* (a liquid that changes viscosity electroactively) (Monkman, 1992), *electrocutaneous stimulators* (that covert visual information into a pattern of vibrations or electrical charges on the skin), *ultrasonic friction displays*, and *rotating disks* for creating slip sensations (Murphy, Webster, & Okamura, 2004).

Sight and Sound

Although the quality of video and audio are commonly measured separately, considerable findings shows that audio and video information is symbiotic in nature, that one medium can have an impact on the user's perception of the other (Rimmel,, Hollier, & Voelcker, 1998; Watson & Sasse, 1996). Moreover, the majority of user multimedia experience is based on both visual and auditory information. As a symbiotic relationship has been demonstrated between the perception of audio and video media, in this chapter we consider multimedia studies concerning the variation and perception of sight and sound together.

Considerable work has been done looking at different aspects of perceived audio and video quality at many different levels. Unfortunately, as a result of multiple influences on user perception of distributed multimedia quality, providing a succinct, yet extensive review of such work is extremely complex.

To this end, we propose an extended version of a model initially suggested by Wikstrand (2003), in which quality is segregated into three discrete levels: the *network-level*, the *media-level* and the *content-level*. Wikstrand showed that all factors influencing distributed multimedia quality (specifically audio and/or video) can be categorised by assessing and categorising the specific information abstraction. The network-level concerns the transfer of data and all quality issues related to the flow of data around the network. The media-level concerns quality issues relating to the transference methods used to convert network data to perceptible media information, that is, the video and audio media. The content-level concerns quality factors that influence how media information is perceived and understood by the end user.

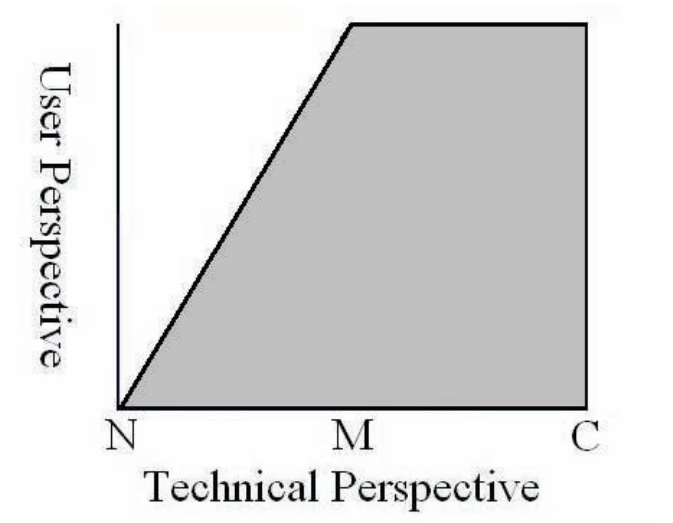
In our work, and in addition to the model proposed by Wikstrand, we incorporated two distinct quality perspectives, which reflect the infotainment duality of multimedia: the user-perspective and the technical-perspective.

- **User-Perspective:** The user-perspective concerns quality issues that rely on user feedback or interaction. This can be varied and measured at the media- and content-levels. The network-level does not facilitate the user-perspective since user perception cannot be measured at this low level abstraction.
- **Technical-Perspective:** The technical-perspective concerns quality issues that relate to the technological factors involved in distributed multimedia. Technical parameters can be varied and measured at all quality abstractions.

At each quality abstraction defined in our model, quality parameters can be varied, for example, jitter at the network-level, frame rate at the media-level, and the display-type at the content-level. Similarly, at each level of the model, quality can be measured, for example, percentage of loss at the network-level, user mean opinion score (MOS) at the media-level, and task performance at the content-level. By determining the abstraction and perspective at which experimental variables are both varied and measured, it is possible to place any multimedia experiment in context of our model.

Using the above model we aid to produce a succinct yet extensive summary of video/ audio multimedia research. The subsequent sections describe this video/audio literature in context of the identified quality definition model (see Figure 1). Each section concerns a quality abstraction level (network, media, or content-level) and includes work relating to studies that adapt and measure quality factors at the defined perspectives (both technical- or media-perspective). Each subsection includes a summary of all relevant studies. A detailed description of new studies is given at the end of the relevant section.

Figure 1. Quality model incorporates network (N), media (M), and content-level (C) abstractions and technical- and user-perspectives dimensions



The Network-Level

Network-Level Quality (Technical-Perspective)

- Claypool and Tanner (1999) manipulated jitter and packet loss to test the impact on user quality opinion scores.
- Ghinea and Thomas (2000) manipulated bit error, segment loss, segment order, delay, and jitter in order to test the impact of different media transport protocols on user perception understanding and satisfaction of a multimedia presentation.
- Procter, Hartswood, McKinlay, and Gallacher (1999) manipulated the network load to provoke degradations in media quality.
- Loss occurs due to network congestion and can therefore be used as an indication of end-to-end network performance or "quality" (Ghinea, Thomas, & Fish, 2000; Koodli & Krishna, 1998).
- A delay is always incurred when sending distributed video packets; however, the delay of consecutive packets is rarely constant. The variation in the period of delay is called the jitter. Wang et al. (2001) used jitter as an objective measure of video quality.
- The end-to-end bandwidth is defined as the network resource that facilitates the provision of these media-level technical parameters. Accordingly, the available end-to-end bandwidth is important since it determines the network resource available to applications at the media-level. Wang, Claypool, and Zuo (2001) measured bandwidth impact for their study of real-world media performance.

Claypool and Tanner (1999): Claypool and Tanner measured and compared the impact of jitter and packet loss on perceptual quality of distributed multimedia video.

Results showed that:

- Jitter can degrade video quality nearly as much as packet loss. Moreover, the presence of even low amounts of jitter or packet loss results in a severe degradation in perceptual quality. Interestingly, higher amounts of jitter and packet loss do not degrade perceptual quality proportionally.
- The perceived quality of low temporal aspect video, that is, video with a small difference between frames, is not impacted as much in the presence of jitter as video with a high temporal aspect.
- There is a strong correlation between the average number of quality degradation events (points on the screen where quality is affected) and the average user quality rating recorded. This suggests that the number of degradation events is a good indicator of whether a user will like a video presentation affected by jitter and packet loss.

Ghinea and Thomas (2000): Ghinea and Thomas tested the impact on user Quality of Perception (QoP) of an adaptable protocol stacks geared towards human requirements (RAP - Reading Adaptable Protocol), in comparison to different media transport protocols (TCP/IP, UDP/IP). RAP incorporates a mapping between QoS parameters (bit error rate, segment loss, segment order, delay and jitter) and Quality of Perception (the user understanding and satisfaction of a video presentation). Video material used included 12- windowed (352*288 pixel) MPEG-1 video clips, each between 26 and 45 seconds long, with a consistent objective sound quality, colour depth (8-bit) and frame rate (15 frames per second). The clips were chosen to cover a broad spectrum of subject matters, while also considering the dynamic, audio, video, and textual content of the video clip.

Results showed that:

- RAP enhanced user understanding, especially if video clips are highly dynamic. TCP/IP can be used for relatively static clips. UDP/IP performs badly, in context of information assimilation.
- Use of RAP successfully improves user satisfaction (10/12 videos). TCP/IP received the lowest associated satisfaction ratings.
- RAP, which incorporates the QoS to QoP mapping (Ghinea et al., 2000), is the only protocol stack used, which was not significantly different to those identified when video were shown on a standalone system. Accordingly, RAP effectively facilitates the provision of user QoP.

Koodli and Krishna (1998): Koodli and Krishna describe a metric called noticeable loss. Noticeable loss captures inter-packet loss patterns and can be used in source server and network buffers to pre-emptively discard packets based on the "distance" to the previous lost packet of the same media stream. Koodli and Krishna found that incorporation of noticeable loss greatly improves the overall QoS, especially in the case of variable bit rate video streams.

Procter et al. (1999): Procter et al. focus on the influence of different media content and network Quality of Service (QoS) variation on a subject's memory of, and comprehension of, the video material. In addition, Procter et al. focus on the impact of degraded visual information on assimilation of non-verbal information.

A simulation network was used to facilitate QoS variation. Two 4-Mbs token rings were connected by a router, so that packets had to pass through the router before being received by the client. Two background traffic generators were used to generate two network conditions: i) no load (with no traffic), and ii) load (with simulated traffic). During the load condition, packet loss was "bursty" in character, varying between zero and one hundred percent, yet had an overall average between 30% and 40%. Audio quality was subsequently dependent on network conditions. Video material used consisted of two diametrically- opposed presentations: a bank's annual report and a dramatized scene of sexual harassment. Two experiments were used:

- The first experiment was designed to investigate the effects of network QoS on a subject's assessment of quality. Two questionnaires were used: The first concerned the subjective evaluation of quality (with scores 1-5 representing very poor and very good respectively) as well as factors that had impaired quality (caption quality, audio quality, video quality, audio/video synchronisation, and gap in transmission); the second questionnaire measured a participants' comprehension of the material based upon their recall of its factual content. Results showed that subjects rated the quality higher in the non-load than in the load condition, with overall impression, ease of understanding, and technical quality being the significantly better quality with no network load. Two factors were found to significantly impair quality in the no-load network condition: video quality and audio/video synchronisation. When a network load was added, audio quality, video quality, audio/video synchronisation, as well as transmission gaps were found to significantly impair user perception of multimedia quality. No difference was measured in the level of factual information assimilated by users.
- The second experiment investigated the impact of visual degradation of the visual channel on the uptake of non-verbal signals. Again, two network load conditions were used: i) no load, and ii) load, to simulate network traffic. The same test approach was used; however the second questionnaire considered: a) factual questions, b) content questions relating to what they thought was happening in a dramatised section of the

video, that is, the participants' ability to judge the emotional state of people, and c) questions asking the user to specify his/her confidence with his/her answers. Results showed that subjects rated quality higher in the no-load condition, with overall impression, content, ease of understanding, and technical quality rating being significantly higher under no-load conditions. In the no-load condition, audio, video quality, and audio-video synchronisation were considered to have an effect on user perception of multimedia quality. In the load network condition, caption quality, audio quality, video quality, audio/video synchronisation, and gap in transmission were all shown to have an impairing impact on user perception of multimedia quality. No significant difference was measured between the level of factual information assimilated by users when using load and non-load conditions. In conclusion, Procter et al. (1999) observed that degradation of QoS has a greater influence on a subject's uptake of emotive/affective content than on their uptake of factual content.

Wang, Claypool, and Zuo (2001): Wang et al. presented a wide-scale empirical study of RealVideo traffic from several Internet servers to many geographically-diverse users. They found that when played over a best effort network, RealVideo has a relatively reasonable level of quality, achieving an average of 10 video frames per second and very smooth playback. Interestingly, very few videos achieve full-motion frame-rates. Wang et al. showed that: With low-bandwidth Internet connection, video performance is most influenced by the user's limited bandwidth connection speed; with high-bandwidth Internet connection, the performance reduction is due to server bottlenecks. Wang et al. (2001) used level of jitter and video frame rates as objective measures of video quality.

The Media-Level

Media-Level Quality (Technical-Perspective)

- Apteker et al. (1995), Ghinea and Thomas (1998), Kawalek (1995), Kies, Williges, and Rosson (1997), Masry, Hemami, Osberger, and Rohaly (2001), Wilson and Sasse (2000a, 2000b), and Wijesekera, Srivastava, Nerode, and Foresti (1999) manipulated video or audio frame rate.
- Ardito, Barbero, Stroppiana, and Visca (1994) developed a metric, which aimed to produce a linear mathematical model between technical and user subjective assessment.
- Gulliver and Ghinea (2003) manipulated use of captions.
- Kies et al. (1997) manipulated image resolution.
- Quaglia and De Martin (2002) used *Peak Signal-to-Noise Ratio* (PSNR) as an objective measure of quality.
- Steinmetz (1996) and Wijesekera et al. (1999) manipulated video skew between audio and video, that is, the synchronisation between two media. In addition, Steinmetz (1996) manipulated the synchronisation of video and pointer skews.
- Teo and Heeger (1994) developed a normalised model of human vision. Extensions to the Teo and Heegar model included: the *Colour Masked Signal to Noise Ratio* metric (CMPSNR) - Van den Branden Lambrecht and Farell (1996), used for measuring the quality of still colour pictures, and the *Normalised Video Fidelity Metric* (NVFM) - Lindh and van den Branden Lambrecht (1996), for use with multimedia video. Extensions of the CMPSNR metric include: the *Moving Picture Activity Metric* (MPAM) - Verscheure and Hubaux (1996), the *Perceptual Visibility Predictor* metric

(PVP) - Verscheure and van den Branden Lambrecht (1997), and the *Moving Pictures Quality Metric* (MPQM) - van den Branden Lambrecht and Verscheure (1996).

- Wang, Claypool, and Zuo (2001) used frame rate as an objective measure of quality. Although the impact of frame rate is adapted by Apteker et al. (1995); Ghinea and Thomas (1998); Kawalek (1995); Kies et al. (1997); Masry et al .(2001); Wijesekera et al (1999), Wang et al. (2001) is, to the best of our knowledge, the only study that used output frame rate as the quality criterion.
- Wikstrand and Eriksson (2002) used various animation techniques to model football matches for use on mobile devices.

Apteker et al. (1995): Apteker et al. defined a *Video Classification Scheme* (VCS) to classify video clips (see Table 1), based on three dimensions, considered inherent in video messages: the temporal (T) nature of the data, the importance of the auditory (A) components and the importance of visual (V) components.

"High temporal data" concerns video with rapid scene changes, such as general sport highlights, "Low temporal data" concerns video that is largely static in nature, such as a talk show. A video from each of the eight categories was shown to users in a windowed multitasking environment. Each multimedia video clip was presented in a randomised order at three different frame rates (15, 10, and 5 frames per second). The users then rated the quality of the multimedia videos on a seven-point graded scale. Apteker et al. showed that:

Table 1. Video classification examples - adapted from Apteker et al (1995)

Category Number	Video Information	*Definition*
1	Logo/ Test Pattern	Tlo Alo Vlo
2	Snooker	Tlo Alo Vhi
3	Talk Show	Tlo Ahi Vlo
4	Stand-up Comedy	Tlo Ahi Vhi
5	Station Break	Thi Alo Vlo
6	Sporting Highlights	Thi Alo Vhi
7	Advertisements	Thi Ahi Vlo
8	Music Clip	Thi Ahi Vhi

- Video clips with a lower video dependence (Vlo) were considered as more watchable than those with a high video dependence (Vhi).
- Video clips with a high level of temporal data (Thi) were rated as being more watchable than those with a low level of temporal data (Tlo).
- Frame-rate reduction itself leads to progressively lower ratings in terms of "watchability".
- There exists a threshold, beyond which no improvement to multimedia quality can be perceived, despite an increase in available bandwidth, which is supported by Fukuda, Wakamiya, Murata, and Miyahara, 1997; Ghinea, 2000; Steinmetz, 1996; van den Branden Lambrecht, 1996).

Apteker et al. expressed human receptivity as a percentage measure, with 100% indicating complete user satisfaction with the multimedia data, and showed that the dependency between human receptivity and the required bandwidth of multimedia clips is non-linear (Apteker et al). In the context of bandwidth-constrained environments, results suggest that a limited reduction in human receptivity facilitates a relatively large reduction in bandwidth requirement (the *asymptotic property* of the VCS curves) (Ghinea, 2000).

Ardito et al. (1994): The RAI Italian Television metric attempts to form a linear objective model from data representing subjective assessments concerning the quality of compressed images (Ardito et al., 1994; Ghinea, 2000). During subjective assessment, the participants were

presented with a sequence of pairs of video clips, one representing the original image and the other showing the degraded (compressed) equivalent. The user is not told which of the two is the original, yet is asked to categorise the quality of the two images using a five- point Likert double-stimulus impairment scale classification similar to the CCIR Rec. 500-3 scale (CCIR, 1974), with scores of one and five representing the "very annoying" and, respectively, "imperceptible" difference between the original and degraded images. All results are then normalised with respect to the original.

The RAI Italian Television metric initially calculates the SNR for all frames of the original and degraded video clips. To enable processing over time (the temporal variable) the SNR values are calculated across all frames, as well as in subsets of specified length *l*. Minimum and maximum values of the SNR are then determined for all groups with *l* frames, thus highlighting noisy sections of video. RAI Italian Television metric considers human visual sensitivity, by making use of a Sobel operator. A Sobel operator uses the luminance signal of surrounding pixels, in a 3x3 matrix, to calculate the gradient in a given direction. Applied both vertically and horizontally, Sobel operators can identify whether or not a specific pixel is part of an edge. The 3x3 matrix returns a level of luminance variation greater or smaller, respectively, than a defined threshold. An "edge image" for a particular frame can be obtained by assigning a logical value of 1 or 0 to each pixel, depending on its value relative to the threshold. An edge image is defined for each frame, facilitating the calculation of SNR for three different scenarios: across the whole frame, only for areas of a frame belonging to edges, and, finally, across the whole frame but only for those areas not belonging to edges. Results show that, the RAI Italian Television linear model can successfully capture 90% of the given subjective information. However, large errors occur if the subjective data is applied across multiple video clips (Ardito et al., 1993), implying high content dependency.

Ghinea and Thomas (1998): To measure the impact of video Quality of Service (QoS) variation on user perception and understanding of multimedia video clips, Ghinea and Thomas presented users with a series of 12 windowed (352x288 pixel) MPEG-1 video clips, each between 26 and 45 seconds long, with a consistent objective sound quality. The clips were chosen to cover a broad spectrum of subject matter, while also considering the dynamic, audio, video, and textual content of the video clip. They varied the frame per second (fps) QoS parameters, while maintaining a constant colour depth, window size, and audio stream quality. Frame rates of 25 fps, 15 fps and 5 fps were used and were varied across the experiment, yet for a specific user they remained constant throughout. Ten users were tested for each frame rate. Users were kept unaware of the frame rate being displayed. To allow dynamic (D), audio (A), video (V) and textual (T) considerations to be taken into account, in both questionnaire design and data analysis, characteristic weightings were used on a scale of 0-2, assigning importance of the inherent characteristics of each video clip.

Table 2. Video characteristics as defined by Ghinea and Thomas (1998)

Video Category	Dynamic (D)	Audio (A)	Video (V)	Text (T)
1 – Commercial	1	2	2	1
2 – Band	1	2	1	0
3 – Chorus	0	2	1	0
4 – Animation	1	1	2	0
5 – Weather	0	2	2	2
6 – Documentary	1	2	2	0
7 – Pop Music	1	2	2	2
8 – News	0	2	2	1
9 – Cooking	0	2	2	0
10 – Rugby	2	1	2	1
11 – Snooker	0	1	1	2
12 – Action	2	1	2	0

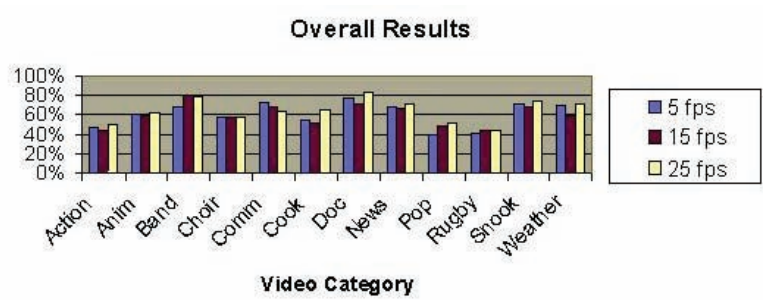
Table 2 contains the characteristic weightings, as defined by Ghinea and Thomas (1998). The clips were chosen to present the majority of individuals with no peak in personal interest, which could skew results. Clips were also chosen to limit the number of individuals watching the clip with previous knowledge and experience.

After the user had been shown a video clip, the video window was closed, and questions were asked about the video that they had just seen. The number of questions was dependent on the video clip being shown and varied between 10 and 12. Once a user had answered all questions relating to the video clip, and all responses had been noted, users were asked to rate the quality of the clip using a six-point Likert scale, with scores one and six representing the worst and, respectively, the best possible perceived level of quality. Users were instructed not to let personal bias towards the subject matter influence their quality rating of the clip. Instead they were asked to judge a clip's quality by the degree to which they, the users, felt satisfied with the service of quality. The questions, used by Ghinea and Thomas, were designed to encompass all aspects of the information being presented in the clips: *D, A, V, T.* A number of questions were used to analyse a user's ability to absorb multiple media at one point in time, as correct answers could only be given if a user had assimilated information from multiple media. Lastly, a number of the questions were used that couldn't be answered by observation of the video alone, but by the users making inference and deductions from the information that had just been presented.

The main conclusions of this work include the following:

- A significant loss of frames (that is, a reduction in the frame rate) does not proportionally reduce the user's understanding and perception of the presentation (see Figure 2). In fact, in some instances the participant seemed to assimilate more information, thereby resulting in more correct answers to questions. Ghinea and Thomas proposed that this was because the user has more time to view a specific frame before the frame changes (at 25 fps, a frame is visible for only 0.04 sec, whereas at 5 fps a frame is visible for 0.2 sec), hence absorbing more information.
- Users have difficulty in absorbing audio, visual, and textual information concurrently. Users tend to focus on one of these media at any one moment, although they may switch between the different media. This implies that critical and important messages in a multimedia presentation should be delivered in only one type of medium.
- When the cause of the annoyance is visible (such as lip synchronisation), users will dis-

Figure 2. Effect of varied QoS on user quality of perception (1998)



regard it and focus on the audio information if considered contextually important.

- Highly dynamic scenes, although expensive in resources, have a negative impact on user understanding and information assimilation. Questions in this category obtained the least number of correct answers. However, the entertainment values of such presentations seem to be consistent, irrespective of the frame rate at which they are shown. The link between entertainment and content understanding is therefore not direct.

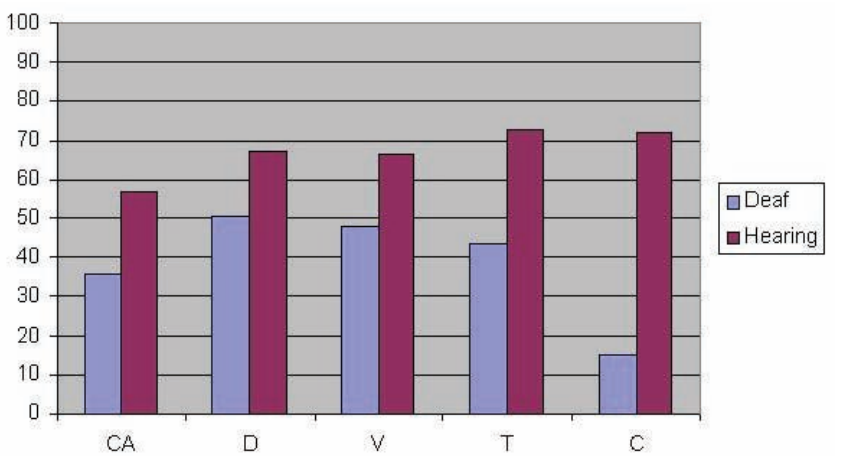
Ghinea and Thomas's method of measuring user perception of multimedia quality, later termed Quality of Perception (QoP), incorporates both a user's capability to understand the informational content of a multimedia video presentation, as well as his/her satisfaction with the quality of the visualised multimedia. QoP has been developed at all quality abstractions of our model (network-, media- and content-level) in cooperation with a number of authors: Fish (Ghinea et al. 2000), Gulliver (Gulliver and Ghinea, 2003; 2004), Magoulas (Ghinea and Magoulas, 2001) and Thomas (Ghinea and Thomas, 1998; 2000; 2001) concerning issues including the development of network protocol stacks, multimedia media assessment, attention tracking, as well user accessibility.

Gulliver and Ghinea (2003): Gulliver and Ghinea used an adapted version of QoP to investigate the impact that hearing level has on user perception of multimedia, with and without captions. They showed users a series of 10 windowed (352x288 pixel) MPEG-1 video clips, each between 26 and 45 seconds long, with a consistent sound quality and video frame rate. Additional captions were added in a separate window, as defined by the experimental design.

Results showed that deafness significantly impacts a user's ability to assimilate information (see Figure 3). Interestingly, use of captions does not increase deaf information assimilation, yet increases quality of context-dependent information assimilated from the caption/audio.

To measure satisfaction, Gulliver and Ghinea (2003) used two 11-point scales (0-10) to measure Level of Enjoyment (QoP-LoE) and user self-predicted level of Information Assimilation (QoP-PIA). A positive correlation was identified between QoP-LoE and QoP-PIA, independent of hearing level or hearing type, showing that a

Figure 3. A detailed breakdown of deaf /hearing information assimilation (%): (CA) caption window / audio, (D) dynamic information, (V) video information,(T) textual information and (C) captions contained in the video window



user's perception concerning their ability to assimilate information is linked to his/her subjective assessment of enjoyment.

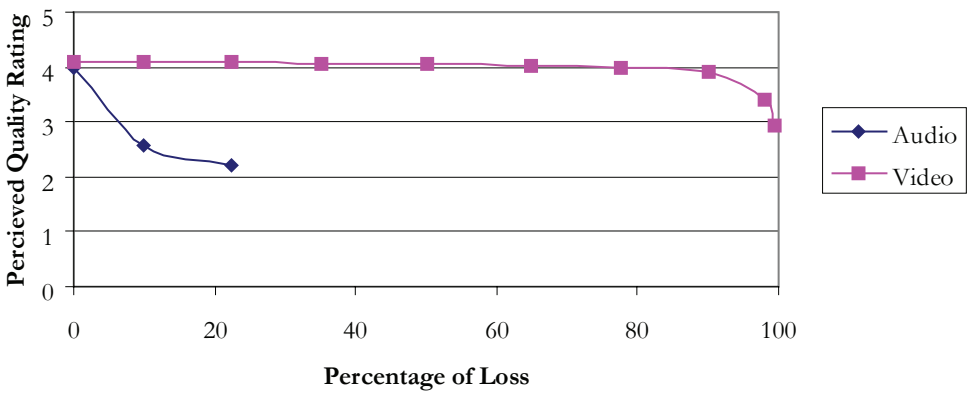
Kawalek (1995): Kawalek showed that loss of audio information has a more noticeable effect on the assimilation of informational content than video frame loss. Users are therefore less likely to notice degradation of video clips if shown low quality audio media (see Figure 4).

Kies et al. (1997): Kies et al. conducted a two-part study to investigate the technical parameters affecting a *Desktop Video Conferencing system* (DVC). Consequently, three frame rate conditions (1, 6 and 30 fps), two resolution conditions (160x120 and 320x240), and three-communication channel conditions were manipulated. Dependent measures included the results of a questionnaire and subjective satisfaction, specifically concerning the video quality. Like Ghinea and Thomas (1998) and Procter et al. (1999), results suggest that factual information assimilation does not suffer under reduced video QoS, but subjective satisfaction is significantly decreased. In addition, a field study was used to look at the suitability of DVC for distance learning. Interestingly, field studies employing similar dependent measures indicated that participants may be more critical of poor video quality in laboratory settings.

Quaglia and De Martin (2002): Quaglia and De Martin describe a technique for delivering "nearly constant" perceptual QoS when transmitting video sequences over IP Networks. On a frame-by-frame basis, allocation of premium packets (those with a higher QoS priority) depends upon on the perceptual importance of each MPEG *macroblock*, the desired level of QoS, and the instantaneous network state. Quaglia and De Martin report to have delivered nearly constant QoS; however, constant reliance on PSNR and use of frame-by-frame analysis raises issues when considering the user perception of multimedia quality.

Steinmetz (1990, 1992, 1996): Distributed multimedia synchronisation comprises both the definition and the establishment of temporal relationships amongst media types. In a multimedia context this definition can be extended such that synchronisation in multimedia systems comprises content, spatial, and temporal relations between media objects. Perceptually, synchronisation of video and textual information or video and image information can be considered as either: *overlay*, which is information that is used in addition to the video information; or *no overlay*, which is information displayed, possibly in another box, to support the current video information. Blakowski

Figure 4. Perceived effect of transmission quality on user perception - Adapted from Kawalek (1995)



and Steinmetz (1995) distinguished two different types of such media objects:

- **Time-dependent media objects:** These are media streams that are characterised by a temporal relation between consecutive component units. If the presentation duration of all components of an object is equal, then it is called a *continuous media object.*
- **Time-independent media objects:** These consist of media such as text and images. Here the semantics of their content does not depend on time-structures.

An example of synchronisation between continuous media would be the synchronisation between the audio and video streams in a multimedia video clip. Multimedia synchronisation can, however, comprise temporal relations between both time-dependent and time-independent media objects. An example of this is a slide presentation show, where the presentation of the slides has to be synchronised with the appropriate units of the audio stream. Previous work on multimedia synchronisation was done in Blakowski and Steinmetz, (1995) and Steinmetz, (1990, 1992, 1996). as well as on topics devoted to device synchronisation requirements. Steinmetz (1990, 1992, 1996) primarily manipulated media skews to measure how lip pointer and non-synchronisation impacted user perception of what is deemed "out of synch" (Steinmetz, 1996). A presentation was considered as being "in synch" when no error was identified, that is, a natural impression. A presentation was considered as being "out of synch" when it is perceived as being artificial, strange, or even annoying. Table 3 summarises the minimal synchronisation errors, proposed by Steinmetz, that were found to be perceptually acceptable between media (see Table 3).

Further experiments incorporating variation in video content, as well as the use of other languages (Spanish, Italian, French, and Swedish) showed no impact on results. Interestingly, Steinmetz did measure variation as a result of the participant group, implying that user experience of manipulating video affects a user's aptitude when noticing multimedia synchronisation errors.

Teo and Heeger (1994): Teo and Heeger present a perceptual distortion measure that predicts

Table 3. Minimal noticeable synchronisation error Steinmetz (1996)

Media		**Mode, Application**	**QoS**
Video	Animation	Correlated	±120ms
	Audio	Lip synchronisation	±80ms
	Image	Overlay	±240ms
		Non-overlay	±500ms
	Text	Overlay	±240ms
		Non-overlay	±500ms
Audio	Animation	Event correlation (e.g. dancing)	±80ms
	Audio	Tightly coupled (stereo)	±11μs
		Loosely coupled (e.g. background music)	±500ms
	Image	Tightly coupled (e.g. music with notes)	±5ms
		Loosely coupled (e.g. slide show)	±500ms
	Text	Text annotation	±240ms
	Pointer	Audio relates to showed item	(-500ms, +750ms)

image integrity based on a model of the human visual system that fits empirical measurements of: 1) the response properties of neurons in the primary visual cortex, and 2) the psychophysics of spatial pattern detection, that is a person's ability to detect a low contrast visual stimuli.

The Teo-Heeger model consists of four steps:

- **a front-end hexagonally sampled quadrature mirror filter transform** function (Simoncellia & Adelson, 1990) that provides an output similar to that of retina, and is similarly tuned to different spatial orientations and frequencies.
- **squaring** to maximise variation.
- **a divisive contrast normalisation mechanism**, to represent the response of a hypothetical neuron in the primary visual cortex.
- **a detection mechanism** (both linear and non-linear) to identify differences (errors) between the encoded image and the original image.

Participants rated images, coded using the Teo and Heeger perceptual distortion measure, as being of considerably better "quality" than images coded with no consideration to the user-perspective. Interestingly, both sets of test images contain similar RMS and PSNR values.

van den Branden Lambrecht and Verscheure (1996): van den Branden Lambrecht and Verscheure present the Moving Picture Quality Metric (MPQM) to address the problem of quality estimation of digital coded video sequences. The MPQM is based on a multi-channel model of the human spatio-temporal vision with parameters defined through interpretation of psychophysical experimental data.

A spatio-temporal filter bank simulates the visual mechanism, which perceptually decomposes video into phenomena such as contrast sensitivity and masking. Perceptual components are then combined, or pooled, to produce a quality rating, by applying a greater summation weighting for areas of higher distortion. The quality rating is then normalised (on a scale from one to five), using a normalised conversion. van den Branden Lambrecht and Verscheure showed MPQM (moving picture quality metric) to model subjective user feedback concerning coded video quality.

Lindh and van den Branden Lambrecht (1996): Lindh and van den Branden Lambrecht introduced the NVFM model (Normalisation Video Fidelity Metric), as an extension of the normalisation model used by Teo and Heeger. The NVFM output accounts for normalisation of the receptive field responses and inter-channel masking in the human visual system and is mapped onto the one to five quality scale on the basis of the vision model used in the MPQM metric.

Lindh and van den Branden Lambrecht compared NVFM with the Moving Picture Quality Metric (MPQM) (van den Branden Lambrecht & Verscheure, 1996). Interestingly, the results of NVFM model are significantly different from the output of the MPQM, as it has a fast increase in user perceived quality in the lower range of bit rate, that is, a slight increase of bandwidth can result in a very significant increase in quality. Interestingly, saturation occurs at roughly the same bit rate for both metrics (approximately 8 Mbit/sec). Lindh and van den Branden Lambrecht proposed NVFM as a better model of the cortical cell responses, compared to the MPQM metric.

van den Branden Lambrecht and Farrell (1996): van den Branden Lambrecht and Farrell introduced a computation metric, which is termed the Colour Masked Signal to Noise Ratio (CMSNR). CMSNR incorporates opponent-colour (i.e. the stimulus of P-ganglion cells), as well as other aspects of human vision involved in spatial vision including: perceptual decomposition (Gabor Filters), masking (by adding weightings to screen areas), as well as the weighted grouping of neuron outputs. van den Branden Lambrecht and Farrell subsequently validated the CMPSNR

metric, using 400 separate images, thus proving the CMPSNR metric as more able to predict user fidelity with a level of accuracy greater than the mean square error.

Wijesekera et al. (1999): A number of mathematical measures of QoS (Quality of Service) models have been proposed (Towsley, 1993; Wijesekera & Stivastava, 1996). Wijesekera & Stivastava (1996) investigated the perceptual tolerance to discontinuity caused by media losses and repetition. Moreover, Wijesekera et al. considered the perceptual impact that varying degrees of synchronisation error have across different streams. Wijesekera et al. followed the methodology of Steinmetz (1996), that is, the manipulation of media skews, to measure stream continuity and synchronisation in the presence of media losses (Wijesekera & Stivastava, 1996) and consequently, quantified human tolerance of transient continuity and synchronisation losses with respect to audio and video media. Wijesekera et al. (1999) found that:

- Viewer discontent with aggregate losses (i.e. the net loss, over a defined duration) gradually increases with the amount of loss, as long as losses are evenly distributed. For other types of loss and synchronisation error, there is a sharp initial rise in user discontent (to a certain value of the defect), after which the level of discontent plateaus.
- When video is shown at 30fps, an average aggregate loss below 17% is imperceptible, between 17% and 23% it is considered tolerated, and above 23% it is considered unacceptable, assuming losses are evenly distributed.
- Losing two or more consecutive video frames is noticed by most users, when video is shown at 30fps. Losing greater than two consecutive video frames does not proportionally impact user perception of video, as a quality rating plateau is reached. Similarly, loss of three or more consecutive audio frames was noticed by most users. Additional consecutive loss of audio frames does not proportionally impact user perception of audio, as a quality rating plateau is reached.
- Humans are not sensitive to video rate variations. Alternatively, humans have a high degree of sensitivity to audio, thus supporting the findings of Kawalek (1995). Wijesekera et al. (1999) suggest that even a 20% rate variation in a newscast-type video does not result in significant user dissatisfaction, whereas a 5% rate variation in audio is noticed by most observers.
- Momentary rate variation in the audio stream, although initially considered as being amusing, was soon considered as annoying. This resulted in participants concentrating more on the audio defect than the audio content.
- An aggregated audio-video synchronisation loss of more that 20% frames was identified. Interestingly, consecutive synchronisation loss of more than three frames is identified by most users, which is consistent with Steinmetz (1996).

Wilson and Sasse (2000a, 200b): Bouch, Wilson, and Sasse (2001) proposed a three-dimensional approach to assessment of audio and video quality in networked multimedia applications: measuring task performance, user satisfaction, and user cost (in terms of physiological impact). Bouch et al. used their approach to provide an integrated framework from which to conduct valid assessment of perceived QoS (Quality of Service). Wilson and Sasse (2000b) used this three-dimensional approach and measured: Blood Volume Pulse (BVP), Heart Rate (HR) and Galvanic Skin Resistance (GSR), to measure the stress caused when inadequate media quality is presented to a participant. Twenty-four participants watched two recorded interviews conducted, using IP video tools, lasting fifteen minutes each. After every five minutes, the quality of the video was changed,

allowing quality variation over time. Audio quality was not varied. Participants therefore saw two interviews with video frame rates of 5-25-5 fps and 25-5-25 fps respectively. While viewing the videos, participants rated the audio/video quality using the QUASS tool (Bouch, Watson, & Sasse, 1998), a *Single Stimulus Continuous Quality* (SSCQE) system where the participant continuously rated quality on an unlabelled scale. Physiological data was taken throughout the experiment. Moreover, to measure whether users perceived any changes in video quality, a questionnaire was also included. Wilson and Sasse showed that the GSR, HR, and BVP data represented significant increases in stress when a video is shown at 5fps in comparison to 25fps. Only 16% of participants noticed a change in frame rate. No correlation was found between stress level and user feedback of perceived quality.

Wilson and Sasse (2000a) showed that subjective and physiological results do not always correlate with each other, which indicates that users cannot consciously evaluate the stress that degraded media quality has placed upon them.

Wikstrand and Eriksson (2002): Wikstrand and Eriksson used animation to identify how alternative rendering techniques impact user perception and acceptance, especially in bandwidth-constrained environments. An animation or model of essential video activity demands that only context-dependent data is transferred to the user and therefore reduces data transfer. Wikstrand and Eriksson performed an experiment to contrast different animations and video coding in terms of their cognitive and emotional effectiveness when viewing a football game on a mobile phone. Results showed that different rendering of the same video content affects the user's understanding and enjoyment of the football match. Participants who preferred video to animations did so because it gave them a better "football feeling", while those who preferred animations had a lower level of football knowledge and thought that animations were best for understanding the game. Wikstrand and Eriksson concluded that more advanced rendering, at the client end, may be used to optimise or blend between emotional and cognitive effectiveness.

Media-Level Quality (User-Perspective)

- Apteker et al. (1995), measured "watchability" (receptivity) as a measure of user satisfaction concerning video quality.
- Ghinea and Thomas (1998) asked respondents to rate the quality of each clip on a seven- point Likert scale.
- Procter et al. (1999) asked subjects to compare the streamed video against the non-degraded original video. Quality was measured by asking participants to consider a number of statements and, using a seven-point Likert-style scale, for example, "the video was just as good as watching a live lecture in the same room" and "the video was just as good as watching a VCR tape on a normal television".
- Steinmetz (1996) used participant annoyance of synchronisation skews as a measure of quality. In both cases only identified errors are considered as being of low quality.
- Wilson and Sasse (2000a; 2000b) used the *Single Stimulus Continuous Quality* QUASS (Bouch et al., 1998) tool to allow the user to continuously rate the audio / video quality, whilst viewing a video presentation.
- Wikstrand and Eriksson (2002) measured user preference concerning the animation rendering technique.

The media-level is concerned with how the media is coded for the transport of information over the network and/or whether the user perceives the video as being of good or bad quality. Accordingly, studies varying media quality, as a direct result of the user, are limited. The best example of quality-related user media variation concerns attentive displays, which manipulate video quality

around a user's point of gaze. Current attentive display techniques were first introduced in McConkie and Rayner (1975) and Saida and Ikeda (1979) and are used in a wide range of applications, including: reading, perception of image and video scenes, virtual reality, computer game animation, art creation and analysis, as well as visual search studies (Baudisch, DeCarlo, Duchowski, & Geisler, 2003; Parkhurst & Niebur, 2002; Wooding, 2002). The perceptual work relating to the use of attentive displays is of limited benefit to the aim of this work and will therefore not be considered in this chapter.

The Content-Level

Content-Level Quality (Technical-Perspective)

- Ghinea and Thomas (1998), Gulliver and Ghinea (2003), Masry et al. (2001), as well as Steinmetz (1996) all varied experimental material to ensure diverse media content.
- Procter et al. (1999) used diametrically opposed presentations: a bank's annual report and a dramatized scene of sexual harassment.
- Steinmetz (1996) used three different views: head, shoulder, and body, which related to the relative proportion of the newsreader shown in the video window.
- Wilson and Sasse (2000a, 2000b) measure participants Blood Volume Pulse (BVP), Heart Rate (HR) and Galvanic Skin Resistance (GSR), to measure for stress as a result of low quality video.

Content-Level Quality (User-Perspective)

- Apteker et al. (1995) measured "watchability" (receptivity) as a measure of user satisfaction concerning video content along temporal, visual, and audio dimensions. Accordingly, "watchability" covers both media- and content-levels.
- Ghinea and Thomas (1998), and Gulliver and Ghinea (2003) used questionnaire feedback to measure a user's ability to assimilate and understand multimedia information.
- Gulliver and Ghinea (2003) asked participants to predict how much information they had assimilated during IA tasks, using scores of 0 to 10 representing "none" and, respectively, "all" of the information that was perceived as being available. Gulliver and Ghinea also measured a participant's level of enjoyment, using scores of 0 to 10 representing "none" and, respectively, "absolute" enjoyment.
- Gulliver and Ghinea (2003) varied participant demographics to measure changes in multimedia perception as a result of deafness, deafness type, and use of captions.
- Procter et al. (1999) used "ease of understanding", "recall", "level of interest", and "level of comprehension" as quality measures.
- Steinmetz (1996) tested videos using a variety of languages (Spanish, Italian, French, and Swedish) in order to check lip synchronisation errors.
- Watson and Sasse (2000) varied peripheral factors, such as volume and type of microphone, to measure in a CSCW environment, the impact on user perception of audio quality.
- Wikstrand and Eriksson (2002) adapted animation rendering techniques, while maintaining important presentation content. Wikstrand and Eriksson showed that animation rendering affects user's understanding and enjoyment of a football match.

Watson and Sasse (2000): Watson and Sasse showed that volume discrepancies, poor quality microphones, and echo have a greater impact on a user's perceived quality of network audio than packet loss.

INCORPORATING THE USER-PERSPECTIVE

Studies have shown that at the:

- **Network-Level:** Technical-perspective network-level variation of bit error, segment loss, delay, and jitter has been used to simulate QoS deterioration. Technical-perspective network-level measurements of loss, delay, and jitter, as well as allocated bandwidth have all been used to measure network level quality performance.
- **Media-Level:** Technical-perspective media-level variation of video and audio frame rate, captions, animation method, inter-stream audio-video quality, image resolution, media stream skews, synchronisation and video compression codecs have been used to vary quality definition. Technical-perspective media-level measurement is generally based on linear and visual quality models, with the exception of who uses output frame rate as the quality criterion. User-perspective media-level variation requires user data feedback and is limited to attentive displays, which manipulate video quality around a user's point of gaze. User-perspective media-level measurement of quality has been used when measuring user "watchability" (receptivity), assessing user rating of video quality, comparing streamed video against the non-degraded original video, as well as for continuous quality assessment and gauging participant annoyance of synchronisation skews.
- **Content-Level:** Technical-perspective content-level variation has been used to vary the content of experimental material as well as the presentation language. Technical-perspective content-level measurement has, to date, only included stress analysis. User-perspective content-level variation has also been used to measure the impact of user demographics, as well as volume and type of microphone on overall perception of multimedia quality. User-perspective content-level measurement has measured "watchability" (receptivity), "ease of understanding", "recall", "level of interest", "level of comprehension", information assimilation, predicted level of information assimilation, and enjoyment.

Table 4. Comparison of user perceptual studies

Study	Participants	Adapted	Measured
Apteker et al. (1995)	60 students	• Frame rate (M) • Video content (C)	• Watchability (M)(C)
Gulliver and Ghinea (2003)	50 participants (30 hearing / 20 deaf)	• Framerate (M) • Captions (M) • Video content (C) • Demographics (C)	• Information assimilation (C) • Satisfaction (C) • Self perceived ability (C)
Procter et al. (1999)	24 participants	• Network load (N) • Video content (C)	• Comprehension (C) • Uptake of non-verbal information (C) • Satisfaction (M)
Wilson and Sasse (2000a; 2000b)	24 participants	• Frame rate (M)	• Galvanic skin resistance (C) • Heart rate (C) • Blood volume pulse (C) • QUASS (M)
Ghinea and Thomas (1998)	30 participants	• Frame rate (M) • Video content (C)	• Information assimilation (C) • Satisfaction (M)

A number of studies have been considered that measure the user-perspective at the content-level (Apteker et al (1995), Ghinea and Gulliver (2003), Ghinea and Thomas (1998), Procter et al. (1999), Wilson and Sasse (2000a; 2000b). These are summarized in Table 4 which:

1. Lists the primary studies that measure the user-perspective at the content-level, stating the number of participants used in each study.
2. Identifies the adapted quality parameters, and defines the quality abstraction at which each parameter was adapted (N = Network-level, M = Media-level, C = Content-level).
3. Provides a list of the measurements taken for each study and the quality level abstraction at which each measurement was taken (N = Network-level, M = Media-level, C = Content-level).

Inclusion of the user-perspective is of paramount importance to the continued uptake and proliferation of multimedia applications since users will not use and pay for applications if they are perceived to be of low quality. Interestingly, all previous studies fail to either measure the infotainment duality of distributed multimedia quality or comprehensively incorporate and understanding the user-perspective. To extensively consider distributed multimedia quality to incorporate and understand the user-perspective, it is essential that, where possible, both technical- and user-perspective parameter variation is made at all quality abstractions of our model, that is, network-level (technical-perspective), media-level (technical- and user-perspective) and content-level (technical- and user-perspective) parameter variation. Furthermore, in order to effectively measure the infotainment duality of multimedia, that is, information transfer and level of satisfaction, the user-perspective must consider both:

- the user's ability to assimilate/understand the informational content of the video {assessing the content-level user-perspective}.
- the user's satisfaction, both measuring the user's satisfaction with the objective QoS settings {assessing the media-level user-perspective}, and also user enjoyment {assessing the content-level user-perspective}.

SUMMARY

In this chapter we set out to consider work relating to each of the multimedia senses - Olfactory (smell), Tactile / Haptic (touch), Visual (sight) and Auditory (sound) - and how this impacts user perception and ultimately user definition of multimedia quality. We proposed an extended model of distributed multimedia, which helped the extensive analysis of current literature concerning video and audio information. We have compared a number of content-level perceptual studies and showed that all previous studies fail to either measure the infotainment duality of distributed multimedia quality or comprehensively incorporate and fully understanding the user-perspective in multimedia quality definition. In conclusion, we show that greater work is needed to fully incorporate and understand the role of the user perspective in multimedia quality definition.

The author believes that a user will not continue paying for a multimedia system or device that they perceive to be of low quality, irrespective of its intrinsic appeal. If commercial multimedia development continues to ignore the user-perspective in preference of other factors, that is, user fascination (i.e. the latest gimmick), then companies ultimately risk alienating the customer. Moreover, by ignoring the user-perspective, future distributed multimedia systems risk ignoring accessibility issues, by excluding access for users with abnormal perceptual requirements.

We have shown that to extensively consider distributed multimedia quality, and to incorporate and understand the user-perspective, it is essential that, where possible, both technical- and user-perspective parameter variation is considered at all quality abstractions of our model, that is, network-level (technical-perspective), media-level (technical- and user-perspective) and content-level (technical- and user-perspective). Furthermore, in order to effectively measure the infotainment duality of multimedia, that is, information transfer and level of satisfaction, the user-perspective must consider both:

- the user's ability to assimilate/understand the informational content of the video.
- the user's satisfaction, both of the objective QoS settings, yet also user enjoyment {assessing the content-level user-perspective}.

If commercial multimedia development effectively considered the user-perspective in combination with technical-perspective quality parameters, then multimedia provision would aspire to facilitate appropriate multimedia, in context of the perceptual, hardware and network criteria of a specific user, thus maximising the user's perception of quality.

Finally, development of quality definition models, as well as user-centric video adaptation techniques (user-perspective personalisation and adaptive media streaming) offers the promise of truly user-defined, accessible multimedia that allows users interaction with multimedia systems on their own perceptual terms. It seems strange that although multimedia applications are produced for the education and/or enjoyment of human viewers, effective development, integration, and consideration of the user-perspective in multimedia systems still has a long way to go….

REFERENCES

Ahumada, A. J., & Null Jr., C. H. (1993). Image quality: A multidimensional problem. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 141-148). Cambridge, MA: MIT Press.

Apteker, R. T., Fisher, J. A., Kisimov, V. S., & Neishlos, H. (1995). Video acceptability and frame rate. *IEEE Multimedia, 2*(3), Fall, 32-40.

Ardito, M., Barbero, M., Stroppiana, M., & Visca, M. (1994, October 26-28). Compression and quality. In L. Chiariglione (Ed.), *Proceedings of the International Workshop on HDTV '94,* Torino, Italy. Springer Verlag.

Baudisch, P., DeCarlo, D., Duchowski, A. T., & Geisler, W. S. (2003). Focusing on the essential: Considering attention in display design. *Communications of the ACM, 46* (3), 60-66.

Blakowski, G., & Steinmetz, R. (1996). A media synchronisation survey: Reference model, specification, and case studies. *IEEE Journal on Selected Areas in Communications, 14*(1), 5-35.

Bouch, A., Watson, A., & Sasse, M. A. (1998). *QUASS—A tool for measuring the subjective quality of real-time multimedia audio and video.* Poster presented at HCI '98, Sheffield, UK.

Bouch, A., Wilson, G., & Sasse, M. A. (2001). A 3-dimensional approach to assessing end-user quality of service. *Proceedings of the London Communications Symposium* (pp. 47-50).

Caldwell, G., & Gosney, C. (1993). Enhanced tactile feedback (tele-taction) using a multi-functional sensory system. *Proceedings of the IEEE International Conference on Robotics and Automation*, Atlanta, GA (pp. 955-960).

Cater, J. P. (1992). The nose have it! *Presence, 1*(4), 493-494.

Cater, J. P. (1994). Smell/taste: Odours. *Virtual Reality*, 1781.

CCIR (1974). Method for the subjective assessment of the quality of television pictures. *13th Plenary Assembly, Recommendation 50*: *Vol. 11* (pp. 65-68).

Claypool, M., & Tanner, J. (1999). The effects of jitter on the perceptual quality of video. *ACM Multimedia'99 (Part 2),* Orlando, FL (pp. 115-118).

Cohn, M. B., Lam, M., & Fearing, R. S. (1992). Tactile feedback for teleoperation. In H. Das (Ed.), *Proceedings of the SPIE Telemanipulator Technology,* Boston (pp. 240-254).

Davide, F., Holmberg, M., & Lundstrom, I. (2001). Virtual olfactory interfaces: Electronic noses and olfactory displays. *Communications through virtual technology: Identity, community, and technology in the Internet age, Chapter 12* (pp. 193-219). Amsterdam: IOS Press.

Dinh, H. Q., Walker, N., Bong, C., Kobayashi, A., & Hodges, L. F. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proceedings of IEEE Virtual Reality* (pp. 222-228).

Fukuda, K., Wakamiya, N., Murata, M., & Miyahara, H. (1997). QoS mopping between user's preference and bandwidth control for video transport. *Proceedings of the 5th International Workshop on QoS (IWQoS),* New York (pp. 291-301).

Ghinea, G. (2000). *Quality of perception - An essential facet of multimedia communications.* Doctoral dissertation, philosophy degree, Department of Computer Science, The University of Reading, UK.

Ghinea, G., & Magoulas, G. (2001). Quality of service for perceptual considerations: An integrated perspective. *IEEE International Conference on Multimedia and Expo, Tokyo* (pp. 752-755).

Ghinea, G., & Thomas, J. P. (1998). QoS impact on user perception and understanding of multimedia video clips. *Proceedings of ACM Multimedia '98,* Bristol UK (pp. 49- 54).

Ghinea, G., & Thomas, J. P. (2000). Impact of protocol stacks on quality of perception. *Proceedings of the IEEE International Conference on Multimedia and Expo,* New York (Vol. 2, pp. 847-850).

Ghinea, G., & Thomas, J. P. (2001). Crossing the man-machine divide: A mapping based on empirical results. *Journal of VLSI Signal Processing, 29*(1/2), 139-147.

Ghinea, G., Thomas, J. P., & Fish, R. S. (2000). Mapping quality of perception to quality of service: The case for a dynamically reconfigurable communication system. *Journal of Intelligent Systems, 10*(5/6), 607-632.

Gulliver, S. R., & Ghinea, G. (2003). How level and type of deafness affects user perception of multimedia video clips. *Universal Access in the Information Society, 2*(4), 374-386.

Gulliver, S. R., & Ghinea, G. (2004). Starts in their eyes: What eye-tracking reveal about multimedia perceptual quality. *IEEE Transaction on System, Man, and Cybernetics, Part A, 34*(4), 472-482.

Hasser, C., & Weisenberger, J. M. (1993). Preliminary evaluation of a shape memory alloy tactile feedback display. In H. Kazerooni, B. D. Adelstein, & J. E. Colgate (Eds.), *Proceedings of the Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems, ASME Winter Annual Meeting,* New Orleans, LA (pp. 73-80).

Heilig, M. L. (1962). Sensorama simulator. U.S.Patent 3,050,870.

Heilig, M. L. (1992). El cine del futuro: The cinema of the future. *Presence, 1*(3), 279-294.

Howe, R. D., Peine, W. J., Kontarinis, D. A., & Son, J. S. (1995). Remote palpation technology. *IEEE Engineering in Medicine and Biology, 14*(3), 318-323.

Ino, S., Shimizu, S., Odagawa, T., Sato, M., Takahashi, M., Izumi, T., & Ifukube, T. (1993). A tactile display for presenting quality of materials by changing the temperature of skin surface. *Proceedings of the Second IEEE International Workshop on Robot and Human Communication*, Tokyo (pp. 220-224).

Kawalek, J. A. (1995). User perspective for QoS management. *Proceedings of the QoS Workshop Aligned with the 3rd Internations Conference on Intelligence in Broadband Services and Network (IS&N 95)*, Crete, Greece.

Kies, J. K., Williges, R. C., & Rosson, M. B. (1997). Evaluating desktop video conferencing for distance learning. *Computers and Education, 28*, 79-91.

Klein, S. A. (1993). Image quality and image compression: A psychophysicist's viewpoint. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 73-88). Cambridge, MA: MIT Press.

Kontarinis, D. A., & Howe, R. D. (1995). Tactile display of vibratory information in teleoperation and virtual environments. *Presence, 4*(4), 387-402.

Koodli, R., & Krishna, C. M. (1998). A loss model for sorting QoS in multimedia applications. *Proceedings of ISCA CATA-98: Thirteenth International Conference on Computers and Their Applications, ISCA*, Cary, NC (pp. 234-237).

Lindh, P., & van den Branden Lambrecht, C. J. (1996). Efficient spatio-temporal decomposition for perceptual processing of video sequences. *Proceedings of the International Conference on Image Processing*, Lausanne, Switzerland (Vol. 3, pp. 331-334), Lausanne, Switzerland.

Martens, J. B., & Kayargadde, V. (1996). Image quality prediction in a multidimensional perceptual space. *Proceedings of the ICIP*, Lausanne, Switzerland (Vol. 1, pp. 877-880).

Masry, M., Hemami, S. S., Osberger, W. M., & Rohaly, A. M. (2001). Subjective quality evaluation of low-bit-rate video: Human vision and electronic imaging VI. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of the SPIE*, Bellingham, WA (pp. 102-113).

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Phychophysics, 17*, 578-586.

Minsky, M., & Lederman, S. J. (1996). Simulated haptic textures: Roughness. *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME International Mechanical Engineering Congress and Exposition, Proceedings of the ASME Dynamic Systems and Control Division*, Atlanta, GA (Vol. 8, pp. 451-458).

Monkman, G. J. (1992). Electrorheological tactile display. *Presence, 1*(2).

Murphy, T. E., Webster, R. J. (3rd), & Okamura, A. M. (2004). Design and performance of a two-dimensional tactile slip display. *EuroHaptics 2004, Technische Universität München*, Munich, Germany.

Parkhurst, D. J., & Niebur, E. (2002). Variable resolution display: A theoretical, practical, and behavioural evaluation. *Human Factors, 44*(4), 611-629.

Procter, R., Hartswood, M., McKinlay, A., & Gallacher, S. (1999). An investigation of the influence of network quality of service on the effectiveness of multimedia communication. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, New York (pp. 160-168).

Quaglia, D., & De Martin, J. C. (2002). Delivery of MPEG video streams with constant perceptual quality of service. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, Lausanne, Switzerland (Vol. 2, pp. 85-88).

Rimmel, A. M., Hollier, M. P., & Voelcker, R. M. (1998). *The influence of cross-modal interaction on audio-visual speech quality perception.* Presented at the 105th AES Convention, San Francisco.

Roufs, J. A. J. (1992). Perceptual image quality: Concept and measurement. *Philips Journal of Resolution, 47*(1), 35-62.

Ryans, M. A., Homer, M. L., Zhou, H., Manatt, K., & Manfreda, A. (2001). Toward a second generation electronic nose at JPL: Sensing film organisation studies. *Proceedings of the International Conference on Environmental Systems'01.*

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception and Psychophysics, 25,* 119-125.

Simoncelli, E. P., & Adelson, E. H. (1990). Nonseparable extensions of quadrature mirror filters to multiple dimensions. *Proceedings of the IEEE Special Issue on Multi-Dimensional Signal Processing, 78*(4), 652-664.

Steinmetz, R. (1990). Synchronisation properties in multimedia systems. *IEEE Journal Select. Areas Communication, 8*(3), 401-412.

Steinmetz, R., (1992). Multimedia synchronisation techniques experiences based on different systems structures. *Proceedings of the IEEE Multimedia Workshop '92,* Monterey, CA.

Steinmetz, R. (1996). Human perception of jitter and media synchronisation. *IEEE Journal on Selected Areas in Communications, 14*(1), 61-72.

Teo, P. C., & Heeger, D. J. (1994). Perceptual image distortion. *Human Vision, Visual Processing and Digital Display V, IS&T / SPIE's Syposium on Electronic Imaging: Science & Technology,* San Jose, CA, 2179 (pp. 127-141).

Towsley, D. (1993). Providing quality of service in packet switched networks. In L. Donatiello & R. Nelson (Eds.), *Performance of computer communications* (pp. 560-586). Berlin; Heidelberg; New York: Springer.

van den Branden Lambrecht, C. J. (1996). Colour moving pictures quality metric. *Proceedings of the ICIP,* Lausanne, Switzerland (Vol. 1, pp. 885-888).

van den Branden Lambrecht, C. J., & Farrell, J. E. (1996). Perceptual quality metric for digitally coded color images. *Proceedings of the VIII European Signal Processing Conference EUSIPCO,* Trieste, Italy (pp. 1175-1178).

van den Branden Lambrecht, C. J., & Verscheure, O. (1996). Perceptual quality measure using a spatio-temporal model of the human visual system. *Proceedings of the SPIE,* San Jose, CA, 2668 (pp. 450-461).

Verscheure, O., & Hubaux, J. P. (1996). *Perceptual video quality and activity metrics: Optimization of video services based on MPEG-2 encoding.* COST 237 Workshop on Multimedia Telecommunications and Applications, Barcelona.

Verscheure, O., & van den Branden Lambrecht, C. J. (1997). Adaptive quantization using a perceptual visibility predictor. *International Conference on Image Processing (ICIP),* Santa Barbara, CA (pp. 298-302).

Wang, Y., Claypool, M., & Zuo, Z. (2001). An empirical study of RealVideo performance across the Internet. *Proceedings of the First ACM SIGCOMM Workshop on Internet Measurement* (pp. 295-309). New York: ACM Press.

Watson, A., & Sasse, M, A. (1996). Evaluating audio and video quality in low cost multimedia conferencing systems. *Interacting with Computers, 8*(3), 255-275.

Watson, A., & Sasse, M. A. (1997). Multimedia conferencing via multicasting: Determining the quality of service required by the end user. *Proceedings of AVSPN '97,* Aberdeen, Scotland (pp. 189-194).

Watson, A., & Sasse, M.A., (2000). The good, the bad, and the muffled: The impact of different degradations on Internet speech. *Proceedings of the 8th ACM International Conference on Multimedia*, Marina Del Rey, CA (pp. 269-302).

Wijesekera, D., & Stivastava, J. (1996). Quality of service (QoS) metrics for continuous media. *Multimedia Tools Applications, 3*(1), 127-166.

Wijesekera, D., Stivastava, J., Nerode, A., & Foresti, M. (1999). Experimental evaluation of loss perception in continuous media. *Multimedia Systems, 7,* 486-499.

Wilson, G. M., & Sasse, M. A. (2000a). Listen to your heart rate: Counting the cost of media quality. In A. M. Paiva (Ed.), *Affective interactions towards a new generation of computer interfaces* (pp. 9-20). Berlin, DE: Springer.

Wilson, G. M., & Sasse, M. A. (2000b). Do users always know what's good for them? Utilising physiological responses to assess media quality. In S. McDonald, Y. Waern, & G. Cockton (Eds.), *Proceedings of HCI 2000: People and Computers XIV - Usability or Else!* Sunderland, UK (pp. 327-339). Springer.

Wikstrand, G., & Eriksson, S. (2002). Football animation for mobile phones. *Proceedings of NordiCHI* (pp. 255-258).

Wilkstrand, G. (2003). *Improving user comprehension and entertainment in wireless streaming media*. Introducing Cognitive Quality of Service, Department of Computer Science.

Wooding, D. S. (2002). Fixation maps: Quantifying eye-movement traces. *Proceedings of the Symposium on ETRA 2002: Eye Tracking Research & Applications Symposium 2002,* New Orleans, LA (pp. 31-36).

This work was previously published in Digital Multimedia Perception and Design, edited by G. Ghinea and S. Y. Chen, pp. 81-109, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.20
Leveraging Digital Multimedia Training for At-Risk Teens

Timothy Shea
University of Massachusetts Dartmouth, USA

Craig Davis
The Learning Community Group, USA

INTRODUCTION

The digital divide exists in poor countries and wealthy countries, the country side and cities, and across age groups. Useful solutions when trying to "bridge" the digital divide should include collaboration with local groups in order to better understand and meet their needs (Eglash, 2004). The most far-reaching examples of these community-oriented, information and communication technology (ICT) products and services result in social and economic impacts beyond just the use of technology—sometimes referred to as community informatics. This paper offers one such solution where an extremely cost—effective, community-based ICT program was successfully piloted in order to improve the computer and digital multimedia literacy of at-risk teenagers, provide job skills, open up new career opportunities, and begin to improve the overall economic capital of the community. While piloted in an inner-city, the program represents a best practice that is equally applicable to a small rural setting or to a regional educational initiative. More specifically, this paper describes the project, the curriculum, and—through the use of a questionnaire and video interviews—the students' experiences taking the class.

BACKGROUND

In July and August of 2002, two organizations combined forces to create a special six-week learning opportunity for 15 at-risk teenagers from Boston, Massachusetts' inner-city neighborhoods. The goal for the class was to learn film and computer-based multimedia skills, employ those skills through working in teams, and develop a video documentary. What distinguishes this educational opportunity from others is how computer technology was so actively intertwined in both creating

and enhancing the educational experience.

During the first half of the six-week program, students learned about film history, utilized computer software to create and edit audio, music, and video tracks, and practiced performing the different roles necessary to create a documentary—producer, director, interviewer, camera-person, and editor. The students spent the last three weeks working in teams applying their newly learned skills toward the creation of the documentary. The group assignment was to create an actual documentary. Within each team, each student chose a specific role to focus upon in order to develop a depth of skills in that area.

In an extremely brief amount of time the students, ages 13 to 18, had the opportunity to gain professional media experience and build confidence in a variety of technical and team skills. In the process, they utilized their full range of learning styles—from visual, to auditory, to kinesthetic—and exercised their critical thinking skills both individually and in a team setting as the hundreds of details in developing and refining a multi-track, multimedia documentary were worked through.

The sponsoring group for the class was ABCD (Action for Boston Community Development), a private, non-profit, human services agency promoting self-help for people and neighborhoods that serves over 100,000 low-income Boston-area residents annually (ABCD, 2003). One of ABCD's programs, SummerWorks, is a summer jobs program for Boston's low-income, at-risk youth that has been in place for 35 years. For the summer of 2002, the SummerWorks program provided over 1,300 inner-city youth with paid, 25 hour a week community-based summer jobs that included mentoring, tutoring, and educational support. "SummerWorks enrollees worked in social service agencies, downtown non-profit and government agencies, museums, day camps, libraries, health centers, hospitals, and more. Enrollees also participated in workshops that provide job readiness and skill-building workshops" (ABCD, 2002). One of the SummerWorks 2002 opportunities was a special pilot program.

Fifteen of the students who were hired for ABCD's SummerWorks program were randomly selected to participate in a special skills-building program where they would create a video documentary of ABCD's 2002 SummerWorks Program. The students had no prior knowledge of what they would be asked to create, came from various locations, and had no prior experience working with one another. Only two of the students had any previous experience creating digital video or digital audio. No academic credit was given for participation in the program.

The LCG (The Learning Community Group) of Boston, based on years of experience in the media production industry, designed and built the hardware and selected and customized the software needed to create a video documentary. TheLCG is a technology research and teaching organization dedicated to technology access and mastery by all people, regardless of age, gender, ethnicity or economic bracket. They develop programs that provide emerging technology instruction in a multitude of diverse settings: public and private schools, homeless shelters, libraries, community centers, government agencies and corporate offices (TheLCG, 2003).

The technology component of the class involved the utilization of The LCG Mobile Media Studio (MMS). The MMS is a professional and portable digital audio, video, and music production studio. The MMS is used to create and deliver material for the Internet, broadcast television, or a host of other CD and DVD media distribution formats. The hardware components included a high-performance digital audio/video workstation as well as high-end audio production equipment, including speakers, microphones, and noise-canceling headsets. The software components included professional-level programs for: creating electronic music, recording and editing professional audio tracks, recording and editing professional video tracks, creating CDs, and

streaming media on the Web. Student support for using the MMS included printed guides and an online support community through forums (TheLCG, 2002).

The class met five days a week, from 9 A.M. until 3 P.M., for six weeks. The course was taught by a master instructor and film producer from TheLCG and assisted by a staff member from ABCD.

THE CURRICULUM

The overall objective of the six-week program was to develop a 25-minute, multimedia documentary about ABCD's 2002 SummerWorks program. Curriculum objectives leading to the overall objective included:

- Study film history
- Comprehend and use film language
- Gain media awareness
- Gain experience in executing every production role on a film or video set
- Develop film and video production skills
- Use digital video editing technologies fluidly
- Hone the art of storytelling
- Develop skills for working in teams

The curriculum was broken down into modules as described in the following sections.

Module 1, Week 1: Objective = Crash Course in Film History / Photography / Cinematography / Film Language

The first week had four major components:

- Description of class/job objectives
- A crash course on film history, photography, cinematography, and film language
- Initial exposure to the cameras and the video editing software
- Team building and interviewing skills

Although the crash course in film and production concepts was considered "too much like high school" by some in the class, they were able to apply the concepts taught in class effectively. One assignment challenged the students to find examples of the concepts on TV. Students accurately identified:

- An Eisenstein montage within a music video
- The rule of thirds being used on a game show
- Joseph Campbell's monomythic arc being followed in an episode of "SpongeBob Squarepants"

After one week, the students were ready to work as a production team to develop their first film short.

Module 2, Week 2 & 3: Cross-Job Training

During weeks 2 and 3, the class was broken evenly into two groups. The students, in essence, became employees. The students rotated from producer, director, interviewer, camera-person, and editor, trying every position at a number of sites around the Boston area. For example:

- Producers and directors contacted the site they visited, set up an arrival time, and scouted the location beforehand to get ideas of how to capture the site
- Camera-people worked on video taping locations and gained experience using the camera
- Editors imported the resulting footage and edited it to music in order to gain experience in using the editing software

Students typically visited one site per day.

Module 3, Weeks 4, 5 & 6: Working on the Final Project: Creating the Documentary of ABCD's Boston SummerWorks 2002.

In week 4, each student ranked which job they were the most confident at and the instructor— based on observation and student preference—formed the production staff that stayed in place until the end of the project. From this point on, the instructor moved from a teacher role to an advisor role. Students decided where they wanted to visit, who to contact and what to ask. They became a fully functional production team.

Weeks 4 and 5 were spent recording footage from 10 different sites. Week 6, the final week, was spent doing post-production work—directors, producers, and interviewers wrote thank-you letters to the sites visited while the camera-people catalogued the tapes and footage. The class voted on their favorite sites and determined an order for the documentary. The segments were then assembled and the final product was shown to various audiences.

The Final Product

The final product was a 26-minute, professional quality, multi-track video documentary of ABCD's 2002 SummerWorks program—produced in only three weeks. The documentary can be viewed from http://www.thelcg.com/research.htm.

The video in its final form, complete with insightful interviews, professional visual composition, succinct story telling, and sophisticated editing is a notable achievement. However, the student's innate capacity and understanding of syncopation elevates the subject matter producing a final product that rivals a professional production. The digital video comprehension that was gained empowers the participants offering a new medium for self-expression.

Assessment of the Student's Work

Several forms of assessment were used to evaluate each student's progress throughout the program:

- Students kept an ongoing journal for recording their thoughts and comments.
- During the third week, a mid-program assessment was conducted. "One-on-one" meetings were conducted between the instructor, the ABCD assistant, and each student. Each student was given formal feedback, to gauge their excitement, dedication and personal investment in the project, as well as allowing them the opportunity to make suggestions and requests. This mid-program assessment was extremely effective in bridging the academic mode with the production studio end.
- A final assessment was conducted through individual video reflections and a round table discussion.The individual video reflections gave the participants the opportunity to reflect on the process. In addition to this, the instructor and the ABCD assistant conducted a round table discussion about the process and the successes and failures that came with it. Students were not easy on themselves either. At this time they pointed out that they wished they could have had more time to perfect the audio in both recording and editing. These assessments were effective due to the ability of the students to make mature, professional, and sometimes poignant suggestions.

Relating the Course Curriculum to the State of Massachusetts' Language Arts Framework

To make the program relevant to the student's middle school and high school education, the course curriculum was designed to fulfill a num-

ber of the specific learning standards for grades seven through 12 as established in the state-wide curriculum frameworks by the Massachusetts Department of Education (Massachusetts Language Arts Framework, 2001). The course curriculum related specifically to the Massachusetts Language Arts Framework, especially for the following standards: media production; analysis of media; discussion; questioning, listening, and contributing; oral presentation; writing; consideration of audience and purpose; and revising. Details of how the program's curriculum relates to the Massachusetts Language Arts Framework are available from the authors.[1]

THE STUDENT EXPERIENCE

Students became very close through this process. By the nature of the project, creating a documentary, each group had to work together as a cohesive whole as well as take responsibility for their own actions and duties. By the end of the program they were more than classmates, they had become a tight-knit family as evidenced by the hugs and tears shared on the last day and their continued attendance beyond their final payday. As the instructor put it, "Throughout the process I witnessed students staying extra hours without pay four out of five days a week, not just because of the equipment but because they were excited to be creating a product, and enjoyed following its progress. Even once there was nothing else to do, students continued to come in at 9 A.M. and stay until 3 because they wanted it to still be a part of their life." All of the students came in at least once after the program was concluded. Six of the 15 students appeared everyday for two weeks after the program concluded.

Data was collected about the student experience through two means. First, a 15-question student evaluation form made up primarily of 1-to-5 Lickert scale questions that provided space for an open-ended explanation for each answer was administered during the last week of class. Secondly, nine of the students were interviewed individually about their class experiences. Segments of those interviews can be found at http://www.thelcg.com/research.htm.

Questionnaire Results

Twelve out of the 15 students in the class completed the questionnaire. A copy of the questionnaire, based upon a validated instrument from ASTD's 2002 Learning Outcomes Report, is available from the authors (ASTD, 2002). The questions used a 5-point Likert scale, where a "1" meant "Strongly Disagree," "2" meant "Disagree," "3" meant "Neither," "4" meant "Agree," and "5" meant "Strongly Agree."

Overall, students were *very satisfied* with the course (Q15), with an average response of *4.4* (Between Agree (4) and Strongly Agree (5)). Interesting results from the other questions include:

Student's Previous Experience

Student's previous computer experience (Q1 & Q2) was quite varied. Average student use of the computer before starting the class clumped into two groups:

- 50% used the computer six or fewer hours per week
- 33% used the computer 21 or more hours per week

Most of the computer use was for the expected—e-mail and chat, writing papers, and surfing the Internet.

Class Organization and Delivery

- **Understanding course objectives:** Students clearly understood the course objectives and felt the course met the objectives (Q5

& 6, Mean of 4.5 and 4.4, respectively). One student's comment, "I knew what I was responsible for," was representative.

- **Teaching effectiveness:** Students felt the instructor's approach to "teaching and presentation of materials made it easy for me to learn" (Mean of 4.4). Students enjoyed the "hands-on" aspect and being able to "get out and do things." Two students mentioned the program "started off like a class," "being taught and told," and how they liked it better once they started using the software.
- **Pace:** Most students felt they had enough time (Q8), but 17% (2 of 12) did not (answered 1 or 2: strongly disagree or disagree). Several comments mentioned learning a lot but wanting more time.
- **Effectiveness:** All but one student answered "Agree" (4) or "Strongly Agree" (5) to the two questions about what they learned: whether they learned something in the class (Q11) and whether they are confident with what they learned after the class was completed (Q12). One student emphasized the point by saying, "It is on my resume." Another said, "I feel like I could teach someone else." Finally, based on the experience of going out and interviewing people, one student now has "more confidence talking to people I don't know."

Impact on Job Skills and Future Job Aspirations

Seventy-five percent answered "Agree" or "Strongly Agree" to whether they see themselves "getting a job where I can use the knowledge and/or skills gained through this course" (Q14). One student specifically mentioned wanting "to work with film when I get older," another "wants to be a producer." In fact, making use of the skills that he learned and honed during the six-week program, one of the students has begun his own business as a wedding videographer. Another student has completely changed her career goals and now wants to become a producer of documentaries, film, and television. Before this summer she was planning on attending a two-year community college. Now, she has already begun investigating film schools in the area and researching their criteria for incoming freshman. Another student submitted the documentary in a competition for an artistic grant. More than 1,000 students applied and he was awarded the artistic grant.

CONCLUSION

Overall, the results are very positive. Put simply, the two immediate objectives of this ICT program were *to inform* and to have the students *perform*. Students needed to quickly learn film and production concepts as well as hands-on skills such as using a video camera and video editing software. TheLCG's MMS, or Mobile Media Studio, provided a field-tested set of hardware and software that is robust and reasonably easy to learn and use. TheLCG's curriculum, tested and refined over a number of years, provided an effective process for high-school-age students (and even younger) to learn the conceptual foundation, the hands-on skills, as well as the communication, team-building, and design skills needed to create a high-quality video documentary within a few weeks. From the results of the end-of-the-class questionnaires and interviews, students were engaged and they enjoyed the many challenges of this course. As one student said, "It was fun and didn't seem like a hard job, just interesting."

In the long term, students took one large step towards succeeding in the 21st century by becoming more literate in both computers and computers' new language, digital multi-media. For the community, economic capital is enhanced through new job skills and career opportunities. Social and cultural capital grows by understanding new ways of expression and new ways to record and distribute history. For example, once people in a

community know how to use digital video and audio they can create their own documentaries and Internet TV stations.

As a result of the successes of this pilot program, ABCD decided to redesign its University High School's computer lab into a full digital media production studio in which every computer is a Mobile Media Studio. In addition, some of the computers now possess professional music recording and DVD authoring capabilities. ABCD and theLCG are also exploring the creation of a dedicated room to be used for regular Internet TV and Radio broadcasts as well as music production.

REFERENCES

ABCD. (2002). Action for Boston Community Development. Retrieved November 26, 2002, from *www.bostonabcd.org/publicinfo/2002/08-13-2002.htm*

ABCD. (2003). Action for Boston community development. Retrieved September 15, 2003, from *www.bostonabcd.org*

ASTD. (2002). The 2002 ASTD learning outcomes report. The American Society of Trainers and Developers. Retrieved November 26, 2002, from *www.astd.org*

Eglash, R. (2004). Community informatics: a two-way to bridge approach. Retrieved July 10, 2004, from http://www.rpi.edu/~eglash/eglash.dir/ci.htm

TheLCG. (2002). The Learning community group. Retrieved November 26, 2002, from *http://www.theLCG.com/services/MMLdemo.swf*

TheLCG. (2003). The Learning community group. Retrieved September 15, 2003, from *www.theLCG.com*

Massachusetts English Language Arts Curriculum Framework. (2001). Massachusetts Department of Education. Retrieved June 10, 2003, from *http://www.doe.mass.edu/frameworks/current.html*

KEY TERMS

Community Informatics: The use of information and computer technologies (ICT) in communities in order to impact communities socially and economically.

Critical Thinking: An active and systematic cognitive strategy to examine, evaluate, and understand complex issues and personal choices, pose provocative questions, correctly frame and then solve problems, and make decisions on the basis of sound reasoning and valid evidence. This competitive edge requires both rigorous analysis and nimble imagination. (Definition based on www.centerforcriticalimpact.com/definitions.htm definition.)

Learning Style: An individual's unique approach to learning based on strengths, weaknesses, and preferences. Though experts do not agree how to categorize learning styles, an example of a categorization system is one that separates learners into auditory learners, visual learners, and kinesthetic (feeling) learners. (Definition based on e-learningguru.com/gloss.htm definition.)

Likert Scale: A rating scale, typically 1 through 5 or 1 through 7, measuring the strength of agreement with a clear statement. Often administered in the form of a questionnaire used to gauge attitudes or reactions. (Definition based on http://www.isixsigma.com/dictionary/ Likert_Scale-588.htm definition.)

Multi-Media: The use of computers to present text, graphics, video, animation, and sound in an integrated way. Long touted as the future revolution in computing, multi-media applications were, until the mid-90's, uncommon due to the expensive hardware required. With increases in

performance and decreases in price, however, multi-media is now commonplace. Current PCs and PC operating systems are both capable and specifically tuned in order to accommodate the rapidly growing demand for multi-media, especially in the consumer market. (Definition based on www.webopedia.com definition.)

Multi-Track: In traditional recording technology, the ability to layer multiple different audio signals at once. In MIDI software, the ability to layer numerous MIDI data streams, including multiple audio tracks, a video track, etc. (Definition based on http://www.cakewalk.com/tips/desktop-glossary.asp definition.)

Streaming Media: The process by which multi-media files (e.g., audio files, video files, and music files) are delivered through the Internet. Such files are often very large, tens or hundreds of megabytes in size.

Syncopation: A style used in order to vary position of the stress on notes so as to avoid regular rhythm. Syncopation is achieved by accenting a weak instead of a strong beat, by putting rests on strong beats, by holding on over strong beats, and by introducing a sudden change of time signature. This style of composition was exploited to fullest capabilities by jazz musicians, often in improvisation. (Definition based on www.geocities.com/BourbonStreet/Delta/4688/ glossary.htm definition.)

ENDNOTE

[1] The authors thank Ruth Joseph for her help with The State of Massachusetts' Language Arts Framework.

This work was previously published in Encyclopedia of Developing Regional Communities with Information and Communication Technology, edited by S. Marshall, W. Taylor, and X. Yu, pp. 475-480, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Section 6

Managerial Impact

This section presents contemporary coverage of the more formal implications of multimedia technologies, more specifically related to the corporate and managerial impact of the core concepts of multimedia, and how these concepts can be applied within organizations. The design and implementation of multimedia advertising is the focus of this section, which provides a how-to guide for multimedia business and commerce. The managerial research provided in this section allows executives and employees alike to understand the role of multimedia technology in business.

Chapter 6.1
Distanced Leadership and Multimedia

Stacey L. Connaughton
Purdue University, USA

INTRODUCTION

At the dawn of the 21st century, more and more organizations in various industries have adopted geographically dispersed work groups and are utilizing advanced technologies to communicate with them (Benson-Armer & Hsieh, 1997; Hymowitz, 1999; Townsend, DeMarie & Hendrickson, 1998; Van Aken, Hop & Post, 1998). This geographical dispersion varies in form. For example, some organizations have adopted "telecommuting," in which members may work at home, on the road and/or at the office (Hymowitz, 1999). Other organizations have created teams that are globally dispersed. A leader located in Palo Alto, California, for example, may be responsible for coordinating employees in Belgium, China and Mexico.

This article examines the role of communication and multimedia in leading people across time and space. To do so, I first note the significance of distanced work relationships; then, outline various conceptualizations of "distance" evident in the literature; next, discuss the role of multimedia in those relationships; and conclude by forecasting future trends. Throughout the article, the term "distanced leadership" is used to refer to leadership in geographically dispersed contexts.

THE PROLIFERATION OF DISTANCED LEADERSHIP

New organizational forms have become increasingly prevalent in recent years. Indeed, many contemporary organizations and teams span time and space. Physical separation of organizational and/or team members is a defining characteristic of virtual organizations and teams (Jarvenpaa & Leidner, 1998; Majchrzak, Rice, King, Malhotra & Ba, 2000; Warkentin, Sayeed & Hightower, 1997; Wiesenfeld, Raghuram & Garud, 1999), geographically dispersed teams (Connaughton & Daly, 2003, 2004a, 2004b; Shockley-Zalabak, 2002), dispersed network organizations (Rosenfeld, Richman & May, 2004) and telework operations (Hylmo & Buzzanell, 2002; Leonardi, Jackson & Marsh, 2004; Scott & Timmerman, 1999). In these forms, the organization or team is constituted in its interaction and formal and

informal networks. By 2005, 20% of the world's work force is expected to work virtually (Prashad, 2003). Indeed, scholars have called on leadership scholarship to "stretch its boundaries to match the elastic nature of global work" (Davis, 2003, p. 48).

Geographical dispersion affords organizations both opportunities and challenges to both business and communication. Table 1 summarizes these issues as they often appear in the literature.

On the one hand, geographically dispersed teams present organizations with many opportunities. They can help organizations maximize productivity and lower costs (Davenport & Pearlson, 1998). And, they can enable organizations to serve international customers and capitalize on globally dispersed talent (Majchrzak, Rice, King, Malhotra & Ba, 2000; Zaccaro & Bader, 2003). Ideally, this geographical dispersion is designed to foster productivity from, and cooperation among, organizational members, just as if they were co-located with one another (see Handy, 1995; Upton & McAfee, 1996).

Yet geographical dispersion also poses some challenges, specifically with regard to leadership. Previous research indicates that (a) a leader's "social presence" may be more difficult to achieve in distanced settings (Kiesler & Sproull, 1992; Warkentin, Sayeed & Hightower, 1997); (b) trust among leaders and team members may be swift yet fleeting (Jarvenpaa, Knoll & Leidner, 1998); (c) members' identification with the team, organization, and leader may be challenged over distance (Connaughton & Daly, 2004b); and (d) communication among leaders and team members may be complicated by diverse ethnic, communication and organizational backgrounds (Cascio, 1999; Cascio & Shurygailo, 2003).

These challenges are put into perspective when one compares what may take place in physically proximate offices to what often happens in distanced work relationships. It has been suggested that co-located office settings provide more opportunities for organizational members to communicate frequently and spontaneously with each other; they allow for potential to interact immediately for troubleshooting; they foster a forum in which to directly access information; and they enable the development and maintenance of relationships (Davenport & Pearlson, 1998). Often, leaders who are co-located with their team members develop and energize relationships with their team through informal as well as formal interaction. In globally dispersed organizations, however, there may be fewer opportunities to informally communicate, leaving some distanced employees feeling isolated from their leaders and from events that take place at the central organization (Van Aken, Hop & Post, 1998; Wiesenfeld, Raghuram & Garud, 1998).

Table 1. Opportunities and challenges of globally dispersed teams

Opportunities	Challenges
•Reduce costs •Serve international customers/clients •Integrate Global Talent•	•Time zone differences •Language differences •Varied communication norms •Limited face-to-ace Contact

CONCEPTUALIZATIONS OF "DISTANCE"

Research on distanced work relationships, including that related to leadership, defines "distance" in different ways. Some scholars examine physical distance, when individuals and leaders are separated by geography (see Antonakis & Atwater, 2002; Kerr & Jermier, 1978). Other scholars investigate social or psychosocial distance, which often refers to perceived differences in status, rank, authority, social standing and power among leaders and followers, all of which may affect the intimacy and social interactions that take place between leaders and followers (see Antonakis & Atwater, 2002; Napier & Ferris, 1993).

Some researchers conceive of physical distance and social distance as related constructs, functioning in a similar manner (see Howell & Hall-Merenda, 1999). Others argue that physical distance and social distance are distinct and should be considered as separate constructs in research. Among them, Antonakis and Atwater (2002) also add a third dimension of distance, perceived interaction frequency, which they define as the perceived degree to which leaders interact with their followers. They propose that physical distance, social distance and perceived interaction frequency are measurable and are separate dimensions, each of which describes an element of "distance" in dispersed work relationships.

Other research examines how individuals perceive distance in geographically dispersed work contexts. For example, in a study of 46 teleworkers in a variety of industries, Leonardi, Jackson and Marsh (2004) argue that these individuals *manage* distance in various ways. The authors conclude that dispersed individuals do not all perceive distance similarly, and that they manipulate the fact that they are geographically distant from others in order to satisfy individual needs. In the authors' words, "… distance is much more than a mere outcome of the use of ICTs; it is rather a tool virtual team members can use to manage their relationships with their coworkers and their organizations" (p. 169).

MULTIMEDIA AND DISTANCED LEADERSHIP

The published work on multimedia, communication technologies and dispersed leadership can be grouped into two broad categories: that which discusses effective practices for using media to forge connections across time and space; and that which addresses key assumptions in previous research, particularly with respect to the perceived necessity of face-to-face interaction and to the impact physical distance has on work relationships.

Effective Practices

Some published work advances effective practices, highlighting various ways that leaders can utilize multiple media to foster connections with distanced employees across time and space. Among the recommendations, researchers have noted: (a) the creation of Web sites, where project managers can post their "lessons learned" and share effective practices with leaders at other sites; (b) the utilization of electronic forums to advertise what "works" in the regions and to propagate those ideas to headquarters and other remote sites; and (c) the development of internal electronic bulletin boards (one devoted to leaders; another devoted to members), where project leaders and team members can ask questions and receive suggestions from other project leaders and members (see Burtha & Connaughton, 2004; Connaughton & Daly, 2003). Majchrzak, Malhotra, Stamps and Lipnack (2004) note that these virtual work spaces should be considered more than "networked drives with shared files" (p. 134). These virtual work spaces must be accessible to everyone at all times, and a place where the team is reminded of its mission, purpose, decisions and future objectives.

In addition to explaining effective practices, existing research advances propositions about which media function particularly well to achieve various leadership objectives. One of these works is based on a series of interviews with distanced leaders about what media they perceive to be effective in executing various leadership functions across time and space (Connaughton & Daly, 2003). Distanced leaders interviewed in this study perceive that face-to-face communication is optimal for achieving objectives, but acknowledge that it is not always possible when employees and team members are dispersed. The research findings suggest that face-to-face communication is best used to set vision, reach policy decisions and begin to build relationships. When face to face is not an option, regularly scheduled telephone calls are most effectively used

to exchange important task-related information, maintain relationships, appraise performance and coordinate teams. And, electronic mail (e-mail) is most effective to exchange technical information, give specific directions, update interested parties and maintain relationships. (For further discussion of technologies and virtual contexts, see Contractor & Eisenberg, 1990; Ferris & Minielli, 2004; Haythornthwaite & Wellman, 1998; Majchrzak, Rice, King, Malhotra & Ba, 2000.)

Addressing Assumptions of Previous Research

Recent work on distanced leadership has begun to carefully consider two assumptions about working in dispersed contexts. Those assumptions are: (a) that face-to-face communication is related to organizational outcomes; and (b) that physical distance necessarily is an impediment to productive and satisfying work relationships.

- **Assumption:** *Face-to-face communication is critical.* It has been argued that, despite the existence of new media, face-to-face communication is still vitally important to achieving organizational outcomes (Cohen & Prusak, 2001). Some scholarship compares experiences of individuals working proximately with one another (and who can communicate face to face) with individuals working apart from one another. For instance, Warkentin, Sayeed and Hightower (1997) found that face-to-face group members perceive greater team cohesion, and more satisfaction with both the group interaction process and group outcomes than did their distanced counterparts. One conclusion that could be drawn from this research is that individuals prefer to work in close proximity to leaders.

  Zack (1994) and Alge, Wiethoff and Klein (2003) found, however, that although initial face-to-face interactions are quite helpful for teamwork, as time goes on and team members come to better understand one another, mediated communication such as e-mail could be used to accomplish tasks. Scholars are also beginning to explore the processes of teams who never meet face to face and yet still function (Bell & Kozlowski, 2002; Davis, 2003). Continued research in this area may challenge the assumption that face-to-face communication is a necessary ingredient of effective distanced work relationships.
- **Assumption:** *Physical distance necessarily challenges work relationships.* Previous research on distanced work relationships assumes that physical distance complicates performance and leader-follower relationships because distance makes it difficult for leaders to engage in relational and task-related behaviors with followers (see Kerr & Jermier, 1978; Olson & Olson, 2000). Often, scholars contrast distanced leadership with proximate leadership, and claim that physical proximity enables more effective communication between leaders and followers (Yagil, 1998).

  However, the perceived *accessibility* of people in the distanced relationship may matter in predicting important outcomes as well (Cascio & Shurygailo, 2003; Napier & Ferris, 1993). Perceived accessibility refers to the distanced employees' perception that they can contact or reach their leader when so desired. Indeed, previous research has suggested that frequent interaction is critical to establishing a feeling of connection across time and space (Antonakis & Atwater, 2002; Connaughton & Daly, 2004b; Leonardi et al., 2004). And, distance may be perceived in positively valenced ways. As Leonardi et al. (2004) have argued, distance may be strategically managed by some distanced leaders and employees to be an *opportunity* rather than a necessary impediment to work relationships.

FUTURE TRENDS

Thousands of companies in diverse industries now have distanced leaders (see Apgar, 1998; Bryan & Fraser, 1999; Hymowitz, 1999; Lipnack & Stamps, 1997; McCune, 1998). And these leaders face the complex task of managing people who are separated from organizational headquarters by time and space.

Future investigations should consider related organizational trends. For instance, does leadership of geographically dispersed ad hoc teams (that are assembled for short-term projects) differ from the type of distanced leadership described here? If so, how? How does one manage contractors and consultants (who may not have loyalty to the organization) from afar? And, given trends in international customer service, how do organizations effectively serve and lead customers from afar?

Future researchers should also continue to develop theoretical models of distanced leadership as well as continue to conduct empirical work on these and other variables. For instance, it will be important to investigate whether actual physical distance *per se* is the most essential defining feature of a dispersed relationship. Instead, perhaps *physical distance* and *access* to leaders and team members function together to affect relationships and outcomes.

Another important issue for both scholars and practitioners is the assumption made by many that distanced teams have more difficulty than face-to-face teams. That presumption warrants empirical testing. The leaders we have talked with in our research have been quite insistent that face-to-face exchanges offer them the optimal medium for communication. None of the leaders interviewed consider mediated technologies as being effective for handling personnel issues, conflicts and relational development. Yet a question arises: Are these responses tied to levels of experience and training with the technologies, generational differences or other factors? It may be that with more experience using various technologies for communication and more perceived expertise with them that people's preference for face-to-face communication for various tasks may diminish. Future research may find that some distanced employees actually prefer mediated communication with their leader.

CONCLUSION

As organizations become more global, as talent becomes more dispersed and as technologies enable people to do far more from afar, distanced leadership and dispersed work relationships will continue to be important to organizations in the 21st century. Given those trends, the issues discussed in this chapter will become ever more critical for scholars and practitioners to consider.

REFERENCES

Alge, B.J., Wiethoff, C., & Klein, H.J. (2003). When does the medium matter? Knowledge-building experiences and opportunities in decision-making teams. *Organizational Behavior and Human Decision Processes, 91,* 26-37.

Antonakis, J., & Atwater, L. (2002). Leader distance: A review and a proposed theory. *Leadership Quarterly, 13,* 673-704.

Apgar, IV, M. (1998). The alternative workplace: Changing where and how people work. *Harvard Business Review, 76*(3), 121-136.

Bell, B.S., & Kozlowski, S.W.J. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27,* 14-49.

Benson-Armer, R., & Hsieh, T. (1997). Teamwork across time and space. *The McKinsey Quarterly, 4,* 18-27.

Bryan, L.L., & Fraser, J.N. (1999). Getting to global. *The McKinsey Quarterly, 4*, 28-37.

Burtha, M., & Connaughton, S.L. (2004). Learning the secrets of long-distance leadership: Eight principles to cultivate effective virtual teams. *Knowledge Management Review, 7*, 24-27.

Cascio, W.F. (1999). Virtual workplaces: Implications for organizational behavior. In C.L. Cooper & D.M. Rousseau (Eds.), *Trends in organizational behavior* (pp. 1-14). Chichester, UK: John Wiley & Sons.

Cascio, W.F. & Shurygailo, S. (2003). E-leadership and virtual teams. *Organizational Dynamics, 31*, 362-376.

Cohen, D., & Prusak, L. (2001). *In good company: How social capital makes organizations work*. Cambridge, MA: Harvard University Press.

Connaughton, S.L., & Daly, J.A. (2003). Long distance leadership: Communicative strategies for leading virtual teams. In D.J. Pauleen (Ed.), *Virtual teams: Projects, protocols, and processes* (pp. 116-144). Hershey, PA: Idea Group Publishing.

Connaughton, S.L., & Daly, J.A. (2004a). Leading from afar: Strategies for effectively leading virtual teams. In S. Godar & S.P. Ferris (Eds.), *Virtual and collaborative teams: Process, technologies & practice* (pp. 49-75). Hershey, PA: Idea Group Publishing.

Connaughton, S.L., & Daly, J.A. (2004b). Leading in geographically dispersed organizations: An empirical study of long distance leadership behaviors from the perspective of individuals being led from afar. *Corporate Communication: An International Journal, 9*(2), 89-103.

Contractor, N.S., & Eisenberg, E.M. (1990). Communication networks and new media in organizations. In J. Fulk & C. Steinfield (Eds.), *Organizations and communication technology* (pp. 143-172). Thousand Oaks, CA: Sage Publications.

Davenport, T.H., & Pearlson, K. (1998). Two cheers for the virtual office. *Sloan Management Review, 39*, 51-65.

Davis, D.D. (2003). The Tao of leadership in virtual teams. *Organizational Dynamics, 33*(1), 47-62.

Ferris, S.P. & Minielli, M.C. (2004). Technology and virtual teams. In S. Godar & S.P. Ferris (Eds.), *Virtual and collaborative teams: Process, technologies & practice* (pp. 193-211). Hershey, PA: Idea Group Publishing.

Handy, C. (1995, May-June). Trust and the virtual organization. *Harvard Business Review,* 40-50.

Haythornthwaite, C., & Wellman, B. (1998). Work, friendship, and media use for information exchange in a networked organization. *Journal of the American Society for Information Science, 49*, 1101-1114.

Howell, J.M., & Hall-Merenda (1999). The ties that bind: The impact of leader-member exchange, transformational and transactional leadership, and distance on predicting follower performance. *Journal of Applied Psychology, 84*, 680-694.

Hylmo, A., & Buzzanell, P.M. (2002). Telecommuting as viewed through cultural lenses: An empirical investigation of the discourses of utopia, identity, and mystery. *Communication Monographs, 69*, 329-356.

Hymowitz, C. (1999, April 6). Remote managers find ways to narrow the distance gap. *The Wall Street Journal*, B1.

Jarvenpaa, S., & Leidner, D.E. (1998). Communication and trust in global virtual teams. *Journal of Computer Mediated Communication, 3*. Online from *www.ascusc.org/jcmc/vol3/issue4/jarvenpaa.html*

Jarvenpaa, S., Knoll, K., & Leidner, D.E. (1998). Is anybody out there? Antecedents of trust in global virtual teams. *Journal of Management Systems, 14*, 29-64.

Kerr, S., & Jermier, J.M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance, 22,* 375-403.

Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes, 52,* 96-123.

Leonardi, P., Jackson, M., & Marsh, N. (2004). The strategic use of 'distance' among virtual team members: A multidimensional communication model. In S.H. Godar & S.P. Ferris (Eds.), *Virtual and collaborative teams: Process, technologies & practice* (pp. 156-173). Hershey, PA: Idea Group Publishing.

Lipnack, J., & Stamps, J. (1997). *Virtual Teams: Reaching across space, time, and organizations with technology.* New York: John Wiley & Sons.

Majchrzak, A., Malhotra, A., Stamps, J., & Lipnack, J. (2004). Can absence make a team grow stronger? *Harvard Business Review, 82*(5), 131-137.

Majchrzak, A., Rice, R.E., King, N., Malhotra, A., & Ba, S. (2000). Technology adaptation: The case of a computer supported inter-organizational virtual team. *MIS Quarterly, 24,* 569-600.

McCune, J.C. (1998). Telecommuting revisited. *Management Review, 87,* 10-16.

Napier, B.J., & Ferris, G.R. (1993). Distance in organizations. *Human Resource Management Review, 3,* 321-357.

Olson, G.M., & Olson, J.S. (2000). Distance matters. *Human Computer Interaction, 15,* 139-178.

Prashad, S. (2003, October 23). Building trust tricky for "virtual" teams. *Toronto Star,* K06.

Rosenfeld, L., Richman, J.M., & May, S.K. (2004). Information adequacy, job satisfaction and organizational culture in a dispersed-network organization. *Journal of Applied Communication Research, 32,* 28-54.

Scott, C.R., & Timmerman, C.E. (1999). Communication technology use and multiple workplace identifications among organizational teleworkers with varied degrees of virtuality. *IEEE Transactions on Professional Communication, 42,* 240-260.

Shockley-Zalabak, P. (2002). Protean places: Teams across time and space. *Journal of Applied Communication Research, 30,* 231-250.

Townsend, A.M., DeMarie, S.M., & Hendrickson, A.R. (1998). Virtual teams: Technology and the workplace of the future. *Academy of Management Executive, 12,* 17-29.

Upton, D.M., & McAfee, A. (1996). The real factory. *Harvard Business Review, 74*(4), 123-133.

Van Aken, J.E., Hop, L., & Post, G.J.J. (1998). The virtual organization: A special mode of strong interorganizational cooperation. In M.A. Hitt, J.E. Ricart I Costa & R.D. Nixon (Eds.), *Managing strategically in an interconnected world* (pp. 301-320). Chichester, UK: John Wiley & Sons.

Warkentin, M.E., Sayeed, L., & Hightower, R. (1997). Virtual teams vs. face-to-face teams: An exploratory study of a Web-based conference system. *Decision Sciences, 28,* 975-996.

Wiesenfeld, B.M., Raghuram, S., & Garud, R. (1999). Communication patterns as determinants of identification in a virtual organization. *Organization Science, 10,* 777-790.

Yagil, D. (1998). Charismatic leadership and organizational hierarchy: Attribution of charisma to close and distant leaders. *Leadership Quarterly, 9,* 161-176.

Zaccaro, S.J., & Bader, P. (2003). E-leadership and the challenges of leading e-teams. *Organizational Dynamics, 31,* 377-387.

Zack, M.H. (1994). Electronic messaging and communication effectiveness in an ongoing work group. *Information and Management, 26*, 231-241.

KEY TERMS

Accessibility: An individual's perception that he/she can contact or reach his/her leader when so desired.

Dispersed/Distributed Teams: Teams separated by some degree of physical distance.

Distanced Leadership: Leadership of a team or organizational members that are separated by some degree of time and distance from their leader.

Identification: Identification is the process in which an individual comes to see an object (e.g., an individual, group, organization) as being definitive of oneself and forms a psychological connection with that object. Although scholars have offered a variety of conceptual definitions for identification, we view it as a communicative process, rooted in discourse and constituting a communicative expression of one's identity.

Social Presence: The perception of physical and/or psychological access to another. Social presence theory often focuses on the aspects of communication media that permit people to connect or "be present" with others and the theory sees some degree of social connectedness as crucial to work relationships.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 226-232, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.2
Leadership Competencies for Managing Global Virtual Teams

Diana J. Wong-MingJi
Eastern Michigan University, USA

INTRODUCTION

The demand for leadership competencies to leverage performance from global virtual teams (GVTs) is growing as organizations continue to search for talent, regardless of location. This means that the work of virtual leaders is embedded in the global shifting of work (Tyran, Tyran & Shepherd, 2003). The phenomenon began with the financial industry as trading took place 24/7 with stock exchanges in different time zones. It is expanding into other industries such as software programming, law, engineering, and call centers. GVTs support the globalization of work by providing organizations with innovative, flexible, and rapid access to human capital. Several forces of competition contribute to the increasing adoption of GVTs, including globalizing of competition, growing service industries, flattening of organizational hierarchies, increasing number of strategic alliances, outsourcing, and growing use of teams (Pawar & Sharifi, 1997; Townsend, DeMarie & Hendrickson, 1998). The backbone of GVTs is innovation with computer-mediated communication systems (CMCSs). Advances with CMCSs facilitate and support virtual team environments.

Leaders of GVTs have a pivotal role in mediating between the internal team processes and the external environment. Leadership competencies also are necessary to keep up with the evolving demands placed on GVTs. Previously, GVTs focused primarily on routine tasks such as data entry and word processing. More recently, the work of GVTs began to encompass non-routine tasks with higher levels of ambiguity and complexity. By tackling more strategic organizational tasks such as launching multinational product, managing strategic alliances, and negotiating mergers and acquisitions, GVTs contribute higher added value to a firm's competitive advantage. As a result, leadership competencies for GVTs become more important in order to maximize the performance of GVTs.

Leadership competencies encompass knowledge, skills, abilities, and behaviors. The following discussion reviews the context, roles, and responsibilities of managing GVTs, identifies five broad categories of GVT leadership competencies, and outlines significant future trends.

BACKGROUND

In order to address specific leadership competencies for GVTs, it is important to understand the virtual workplace context. "Global virtual teams being a novel organizational design, it is very important to maximize the fit between team design and their stated intent" (Prasad & Akhilesh, 2002, p. 104). Currently, many organizations are deploying the use of GVTs much more rapidly than the collective understanding of their unique characteristics, dynamics, and processes. Anecdotal evidence exists about the difficulties and poor performance of GVTs. But the expectations of flexibility, accessing expertise regardless of geographical location, and speed of fulfilling organizational goals continue to drive the growth of GVTs (Gibson & Cohen, 2003).

GVTs have similarities and differences when compared with traditional teams (Maznevski & Chudoba, 2000). The similarities include being guided by shared goals, working on interdependent tasks, and sharing responsibilities for outcomes. The differences are the collocation and synchronous communication of traditional teams vs. geographical dispersion and often asynchronous communication for virtual teams. The stability of GVTs depends on the project and the team's role in fulfilling the organizational purpose. Thus, GVT leaders may be working with a project orientation or indefinite perpetual organizational responsibilities, which shape the lifecycle of the team.

Effective GVT leaders must manage magnified ambiguities and complexities compared to traditional team leaders. Prasad and Akhilesh (2002) define a GVT as "a team with distributed expertise and that spans across boundaries of time, geography, nationality, and culture" (p. 103). They address a specific organizational goal with enhanced performance and operate with very little face-to-face interaction and predominantly computer mediated and electronic communication. As a result, leaders of GVTs need to address unique challenges that stem from spatial distances, asynchronous communication, multicultural dynamics, and national boundaries in a virtual environment.

Established research findings on teams indicates that leaders have a critical influence on team performance outcomes (Bell & Kozlowski, 2002; Fjermestad & Hiltz, 1998-1999; Kayworth & Leidner, 2001-2002). In general, team leaders have two critical functions: team development and performance management. Some general leadership tasks for managing teams include developer of team processes, facilitators of communications, and final arbiter for task completion (Duarte & Tennant-Snyder, 1999). Bell & Kozlowski (2002) offer a typology of virtual teams based on four characteristics—temporal distribution, boundary spanning, lifecycle, and member roles—that are mediated by task complexity. These characteristics imply that effective management of GVTs requires a portfolio of leadership competencies to address the following responsibilities: (1) provide clear direction, goals, structures, and norms to enable self regulation among team members; (2) anticipate problems; (3) monitor the environment and communicate changes to inform team members; (4) design back-up plans to buffer changes in environmental conditions; (5) develop feedback opportunities into team management structure for regular performance updates; (6) diagnose and develop appropriate team development through a virtual medium; (7) diagnose the translation of self-regulation methods across different boundaries; (8) modify behaviors and actions according to the particular situations to support the communication of worldviews among team members and build a third culture; and (9) identify and communicate team member roles to create role networks.

An important component of the GVT leader's work environment is the virtual "rooms" for the team's interactions. A wide range of products offers differing capabilities. For example, Groove Client 2.5 and Enterprise Management from

Groove Networks, Workgroup Suite 3.1 from iCohere, and eRoom 7.0 from Documentum are products that facilitate how virtual teams can navigate through cyberspace (Perey & Berkley, 2003). Large firms in the auto industry use a commercial B2B product called ipTeam from NexPrise to support collaboration among geographically dispersed engineering team members. IBM offers the IBM Lotus Workplace Team Collaboration 2.0. Free Internet downloads such as NetMeeting from Microsoft also are available to facilitate virtual meetings. Competitors include FarSite from DataBeam Corp, Atrium from VocalTec Communications Ltd., ProShare from Intel Corp, and Conference from Netscape. The list of available CMCS products continues to grow and improve with more features that attempt to simulate face-to-face advantages. As a result, part of managing GVTs includes evaluating, selecting, and applying the most appropriate CMCS innovations to support team interactions. Adopting CMCS needs to account for work locations, members involved, technological standardization, work pace, work processes, and nature of work in the organization. In sum, a GVT leadership portfolio must be able to manage CMCSs, diverse team members, team development, and work flow processes.

GVT LEADERSHIP COMPETENCIES

Competencies for GVT leaders can be classified into five broad categories: CMCS proficiency, work process design, cross-cultural competencies, interpersonal communication, and self-management. The five groups of competencies are interrelated. For example, a high degree of expertise with CMCSs without the necessary interpersonal communication competencies likely will lead to conflicts, absentees, and negative productivity.

First, GVT leaders need to have technical proficiency with innovations in CMCS in order to align the most appropriate technological capabilities with organizational needs. Technical knowledge of CMCSs and organizational experience enables GVT leaders to align technology with strategic organizational goals. Organizational experience provides GVT leaders with insights regarding the organizational work task requirements, strategic direction, and culture. This tacit knowledge is rarely codified and difficult to outsource compared to explicit knowledge. This implies that firms should provide training and professional development for leaders to increase CMCS proficiency.

Second, GVT leaders require work process design competencies to manage the workflows. Managing global virtual workflows depends on leadership skills to structure teams appropriately for subtasks, monitor work progress, establish expectations, maintain accountability, build a cohesive team, motivate team members, create trust, develop team identity, and manage conflicts (Montoya-Weiss, Massey & Song, 2001; Pauleen & Yoong, 2001; Piccoli & Ives, 2003). GVT leaders also need to devote considerable attention to performance management, especially in prototypical teams where there may be information delays and members are decoupled from events. GVT leaders can employ temporal coordination mechanisms to mitigate negative effects of avoidance and compromise in conflict management behavior on performance (Montoya-Weiss, Massey & Song, 2001). During the launching of teams, GVT leaders need to use appropriate team building techniques (e.g., discussion forums) to become acquainted and to establish positive relationships (Ahuja & Galvin, 2003; Prasad & Akhilesh, 2002). The lifecycle of virtual teams tends to proceed through four stages of group development that entails forming with unbridled optimism, storming with reality shock, norming with refocus and recommitment, and performing with a dash to the finish (Furst et al., 2004). The lifecycle of virtual teams influences the development of team spirit and identity, which is more important with continuous virtual team lifecycle. Its membership is relatively more stable compared to temporary projects. Task com-

plexity places constraints on team structure and processes (Prasad & Akhilesh, 2002). Relatively simple tasks have less need for stable internal and external linkages, common procedures, and fixed membership, compared to more complex tasks. Leaders need to assert flexible, collegial authority over tasks and act as empathetic mentors to create collaborative connections between team members (Kayworth & Leidner, 2001). In sum, managing the work process design requires dealing with paradoxes and contradictions to integrate work design and team development.

Third, GVT leaders also require cross-cultural competencies, more specifically identified as global leadership competencies. "Successful virtual team facilitators must be able to manage the whole spectrum of communication strategies as well as human and social processes and perform these tasks across organizational and cultural boundaries via new [information and communication technologies]" (Pauleen & Yoong, 2001, p. 205). Developing global leadership competencies entail a sequence from ignorance, awareness, understanding, appreciation, and acceptance/ internalization to transformation (Chin, Gu & Tubbs, 2001). The latter stages involve development of relational competence to become more open, respectful, and self-aware (Clark & Matze, 1999). Understanding cultural differences helps to bridge gaps in miscommunication. Identifying similarities provides a basis for establishing common grounds and interpersonal connections among team members. Leaders who are effective in leading across different cultures have relational competence to build common grounds and trust in relationships (Black & Gregersen, 1999; Gregersen, Morrison & Black, 1998; Manning, 2003). By increasing trust, leaders can connect emotionally with people from different backgrounds to create mutually enhancing relationships (Holton, 2001; Jarvenpaa & Leidner, 1999). The connections are critical to construct a high-performing team (Pauleen, 2003). A key to cross-cultural leadership competencies for GVTs is projecting them into a virtual environment. This is related to CMCS proficiency, which supports the communication cross-cultural competencies in a virtual environment. Cross-cultural competencies also are closely interrelated with both interpersonal communication competencies and self-management to effectively lead GVTs.

Fourth, interpersonal communication competencies do not necessarily encompass cross-cultural competencies. But cross-cultural competencies build upon interpersonal communication competencies. Strong interpersonal communication enables GVT leaders to span multiple boundaries to sustain team relationships (Pauleen, 2003). An important communication practice is balancing the temporal dimension and rhythm of work to stay connected (Maznevski & Chudoba, 2000; Saunders, Van Slyke & Vogel, 2004). Interpersonal communication competencies for GVT leaders need to focus on the human dimension. For example, GVT leaders need to be conscious of how they "speak," listen, and behave non-verbally from their receiver's perspective without the advantage of in-the-moment, face-to-face cues. This provides the basis for moving from low to higher levels of communication—cliché conversation, reporting of facts about others, sharing ideas and judgments, exchanging feelings and emotions, and peak communication with absolute openness and honesty (Verderber & Verderber, 2003). Interpersonal communication skills of GVT leaders should, at a minimum, support the exchange of ideas and judgments. When GVT leaders demonstrate "active listening" online, team members likely will move toward higher levels of communication. Active listening in GVTs can be demonstrated with paraphrasing, summarizing, thoughtful wording, avoiding judgment, asking probing questions, inviting informal reports of progress, and conveying positive respectful acknowledgements. Another aspect of interpersonal communication competencies for GVT leaders is establishing netiquette, which establishes ground rules and team culture. GVT leaders can strategi-

cally develop their interpersonal communication competencies to socialize team members, build team connections, motivate team commitment, resolve conflicts, and create a productive team culture to achieve high performance outcomes (Ahuja & Galvin, 2003; Kayworth & Leidner, 2001-2002).

Finally, a GVT leader's self-management competencies fundamentally influence the development of the four competencies. GVT leaders need to manage their self-assessment and development to acquire a portfolio of competencies. A high level of emotional intelligence enables GVT leaders to engage in self-directed learning for personal and professional development. Self-management refers to adaptability in dealing with changes, emotional self-control, initiative for action, achievement orientation, trustworthiness, and integrity with consistency among values, emotions and behavior, optimistic view, and social competence (Boyatzis & Van Oosten, 2003). The development of GVT leaders with self-management can positively influence team performance by rectifying areas of their own weaknesses and reinforcing their strengths.

In summary, GVTs provide organizations with an important forum for accomplishing work and gaining a competitive advantage in global business. Technological innovations in CMCSs provide increasingly effective virtual environments for team interactions. A critical issue focuses on the GVT leader with the necessary portfolio of competencies. Research and understanding of leadership competencies for managing GVTs are at a nascent stage of development.

FUTURE TRENDS

Researchers need to delve into this organizational phenomenon to advance best practices for multiple constituents and help resolve existing difficulties with GVTs. Understanding leadership competencies for managing GVTs depends on a tighter coupling in the practice-research-practice cycle. Given turbulent competitive environments and more knowledge-based competition, research practices need to keep up with the rapid pace of change. At least three important trends about GVTs need to be addressed in the future.

First, GVTs will continue to grow in strategic importance. An important implication is that GVTs will face greater complexities and ambiguities. Furthermore, GVT leaders will have little or no contextual experience with their team members' locations. This is a significant shift when globe-trotting managers often have face-to-face time with their team members in different locations. Thus, the need to create authentic emotional connections and accomplish the task at hand through multiple CMCSs will continue to be important

Second, another important trend is the rapid pace of technological innovations in telecommunications. New developments will create more future opportunities. For example, advances with media-rich technologies enable communication that narrows the gap between virtual and face-to-face interactions. However, there is little understanding about the relationship between technological adoption and team members from different cultural backgrounds. Given cultural differences, an important consideration would be how people will relate to technological innovations. This has implications for how leaders will manage GVTs. This research issue also has implications for firms engaged in developing CMCSs, because it will affect market adoption.

Last, although not least, organizations also will need to keep pace with the growth of GVTs by developing supporting policies, compensation schemes, and investments. GVT leaders can make important contributions to facilitate organizational development and change management.

The existing GVT literature has some preliminary theoretical developments that require rigorous empirical research. Future research needs to draw from intercultural management,

organization development (OD), and CMCSs with interdisciplinary research teams. OD researchers and practitioners will provide an important contribution to different levels of change—individual, groups and teams, organizational, and interorganizational—as managers and organizations engage in change processes to incorporate GVTs for future strategic tasks.

CONCLUSION

The use of global virtual teams is a relatively new organizational design. GVTs allow organizations to span time, space, and organizational and national boundaries. But many organizational GVT practices have a trial and error approach that entails high costs and falls short of fulfilling expectations. The cost of establishing GVTs and their lackluster performance creates a demand for researchers to figure out how to resolve a range of complex issues. An important starting point is with the leadership for managing GVTs. Developing a balanced portfolio of five major leadership competencies—CMCS proficiency, work process and team designs, cross-cultural competence, interpersonal communication, and self-management—increases the likelihood of achieving high performance by GVTs.

REFERENCES

Ahuja, M.K., & Galvin, J.E. (2003). Socialization in virtual groups. *Journal of Management, 29*(3), 161-185.

Bell, B.S., & Kozlowski, S.W. (2002). A typology of virtual teams: Implications for effective leadership. *Group and Organization Management, 27*(1), 14-49.

Black, J.S., & Gregersen, H.B. (1999). The right way to manage expats. *Harvard Business Review, 77*(2), 52-59.

Boyatzis, R., & Van Oosten, E. (2003). A leadership imperative: Building the emotionally intelligent organization. *Ivey Business Journal, 67*(2), 1-6.

Bueno, C.M., & Tubbs, S.L. (2004). Identifying global leadership competencies: An exploratory study. *Journal of American Academy of Business, 5*(1/2), 80-87.

Chin, C., Gu, J., & Tubbs, S. (2001). Developing global leadership competencies. *Journal of Leadership Studies, 7*(3), 20-31.

Clark, B.D., & Matze, M.G. (1999). A core of global leadership: Relational competence. *Advances in Global Leadership, 1*, 127-161.

Duarte, N., & Tennant-Snyder, N. (1999). *Mastering virtual teams: Strategies, tools, and techniques that succeed.* San Francisco, CA: Jossey-Bass.

Fjermestad, J., & Hiltz, S.R. (1998-1999). An assessment of group support systems experiment research: Methodology and results. *Journal of Management Information Systems, 15*(3), 7-149.

Furst, S.A., Reeves, M., Rosen, B., & Blackburn, R.S. (2004). Managing the life cycle of virtual teams. *Academy of Management Executive, 18*(2), 6-20.

Gibson, C.B., & Cohen, C.B. (Eds.) (2003). *Virtual teams that work: Creating conditions for virtual team effectiveness.* San Francisco, CA: Jossey-Bass.

Gregersen, H.B., Morrison, A.J., & Black, J.S. (1998). Developing leaders for the global frontier. *Sloan Management Review, 40*(1), 21-33.

Holton, J.A. (2001). Building trust and collaboration in a virtual team. *Team Performance Management, 7*(3/4), 36-47.

Jarvenpaa, S.L., & Leidner, D.E. (1999). Communication and trust in global virtual teams. *Organization Science, 10*(6), 791-815.

Kayworth, T.R., & Leidner, D.E. (2001-2002). Leadership effectiveness in global virtual teams. *Journal of Management Information Systems, 18*(3), 7-31.

Manning, T.T. (2003). Leadership across cultures: Attachment style influences. *Journal of Leadership and Organizational Studies, 9*(3), 20-26.

Maznevski, M.L., & Chudoba, K.M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science, 11*(5), 473-492.

Montoya-Weiss, M.M., Massey, A.P., & Song, M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal, 44*(6), 1251-1262.

Pauleen, D.J. (2003). Leadership in a global virtual team: An action learning approach. *Leadership & Organization Development Journal, 24*(3), 153-162.

Pauleen, D.J., & Yoong, P. (2001). Relationship building and the use of ICT in boundary-crossing virtual teams: A facilitator's perspective. *Journal of Information Technology, 16*, 205-220.

Pawar, K.S., & Sharifi, S. (1997). Physical or virtual team collocation: Does it matter? *International Journal of Production Economics, 52*, 283-290.

Perey, C., & Berkley, T. (2003). Working together in virtual facilities. *Network World, 20*(3), 35-37.

Piccoli, G., & Ives, B. (2003). Trust and unintended effects of behavior control in virtual teams. *MIS Quarterly, 27*(3), 365-395.

Prasad, K., & Akhilesh, K.B. (2002). Global virtual teams: What impacts their design and performance? *Team Performance Management, 8*(5/6), 102-112.

Saunders, C., Van Slyke, C., & Vogel, D. (2004). My time or yours? Managing time visions in global virtual teams. *Academy of Management Executive, 18*(1), 19-31.

Townsend, A., DeMarie, S., & Hendrickson, A. (1998). Virtual teams: Technology and the workplace of the future. *Academy of Management Executive, 12*(3), 17-29.

Tyran, K.L., Tyran, C.K., & Shepherd, M. (2003). Exploring emerging leadership in virtual teams. In C.B. Gibson, & C.B. Cohen (Eds.), *Virtual teams that work: Creating conditions for virtual team effectiveness*, (pp. 183-195). San Francisco: Jossey-Bass.

Verderber, R.F., & Verderber, K.S. (2003). *Interact: Using interpersonal communication skills.* Belmont, CA: Wadsworth.

KEY TERMS

Asynchronous Communication: Information exchanges sent and received at different times, often taking place in geographically dispersed locations and time zones.

CMCS: Computer-mediated communication system includes a wide range of telecommunication equipment such as phones, intranets, Internets, e-mail, group support systems, automated workflow, electronic voting, audio/video/data/desktop video conferencing systems, bulletin boards, electronic whiteboards, wireless technologies, and so forth to connect, support, and facilitate work processes among team members.

Colocation: Team members sharing the same physical location, which allows for face-to-face interaction.

Emotional Intelligence: A set of competencies that derive from a neural circuitry emanating in the limbic system. Personal competencies related

to outstanding leadership include self-awareness, self-confidence, self-management, adaptability, emotional self-control, initiative, achievement orientation, trustworthiness, and optimism. Social competencies include social awareness, empathy, service orientation, and organizational awareness. Relationship management competencies include inspirational leadership, development of others, change catalyst, conflict management, influence, teamwork, and collaboration.

Human Capital: The knowledge, skills, abilities, and experiences of employees that provide value-added contributions for a competitive advantage in organizations.

Netiquette: This is a combination of the words "etiquette" and "Internet" ("net," for short). Netiquette is rules of courtesy expected in virtual communications to support constructive interpersonal relationships in a virtual environment.

Synchronous Communication: Information exchanges taking place in the same space and time, often face-to-face.

Temporal Coordination Mechanism: A process structure imposed to intervene and direct the pattern, timing, and content of communication in a group.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 519-525, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.3
Short Message Service (SMS) as an Advertising Medium

Shintaro Okazaki
Autonomous University of Madrid, Spain

INTRODUCTION

The proliferation of the Internet-enabled mobile device has extended into many parts of the world. Collectively, the mobile-network operators paid more than $100 billion for licenses to operate "third-generation" (3G) networks, which were among "the largest bet in business history on the introduction of a new technology" (Economist, 2005). This drastic move has been most illustrated by the use of short message service (SMS) and multimedia messaging service (MMS) by mobile users. For example, a recent survey indicates that SMS in the Asia-Pacific region will increase to up to 75% of mobile subscribers in 2006 (IDC Asia/Pacific, 2003). As a result, marketers and agencies are increasingly interested in taking advantage of this growth, by incorporating SMS advertising as part of an integrated marketing communications (IMC) strategy. However, there has been little academic research on mobile advertising, perhaps because its growth is still in an early stage and the technological infrastructure varies across markets. The study has two objectives: (1) to identify the factors influencing MNCs' managerial intention to adopt SMS advertising, and (2) to test a statistical relationship between these factors and managerial intention to use SMS advertising. To this end, we conducted telephone interviews of senior executives of MNCs operating in European markets.

CONCEPTUAL FRAMEWORK AND HYPOTHESES

Branding Technique

In an environment where building the brand is a fundamental goal for many managers, the need to build brand equity is likely to be at the center of many marketing decisions. Firms using SMS-based campaigns can attract consumer attention and produce consumer responses to a much greater degree than via other direct marketing channels, because SMS has been claimed to be an effective tool in building and testing customer loyalty by developing demographic databases (Mylonopoulos & Doukidis, 2003). From an industry perspective, McDonald's conducted a

text-messaging campaign in conjunction with a popular TV song contest in the UK, offering concert tickets and backstage passes, while entry in the Coca-Cola Grand Sweepstakes Competition was offered to U.S. college students who sent a text message to a number printed on a Diet Coke can (Dano, 2002).

Facilitating Conditions

Lu, Yu, Liu, and Yao (2003) suggest that facilitating conditions are one of the most important determinants, along with the ease of using wireless Internet. In this light, the integration of competing standards and fragmented systems across countries, cross-network support for SMS, and higher connection speeds are all necessary conditions for a wider transmission of mobile advertising. In addition, the availability of Web-enabled mobile handsets with 2.5G or 3G functionality would significantly affect the adoption of MMS-based (multimedia message services) campaigns. In this light, a wider selection of handsets must be available, to enable consumers to choose their preferred combination of necessary functions and diverse features.

Location-Based Services

The satellite-based global positioning system (GPS) offers the ability to tailor services and promotional offers to individual consumers' needs, by locating their position (Sadeh, 2002). Mobile handset makers and content providers are increasingly attracted by the commercial feasibility of applying GPS to their service. For example, on an extended menu of i-mode, "i-area" includes a diverse range of location-based services: weather news, restaurant guide, local hotel information, zoomable maps with an address finder function, and traffic updates and estimation of travel times. This facility would give MNCs strategic leverage in mobile marketing, because individuals' behavior and receptiveness to advertising is likely to be influenced by their location and time, and marketers can thus induce impulse buying by providing the right information for the right place (Barnes, 2002).

Connection Costs

Another important factor is the concept of connection costs. For example, to send or receive one megabyte of data on 2.5G i-mode costs 32 euros (0.3 yen) per packet. At a rate of 19 euro cents per 160-character SMS message, European consumers would have to pay 1,356.98 euros to send one megabyte of data by SMS, or approximately 62 times as much as the Japanese pay (Scuka, 2003). In addition, European mobile operators have passed on to consumers the additional costs incurred in obtaining 3G spectrum licenses, and this has made any dramatic price reduction impossible (Baldi & Thaung, 2002). Such cost factors adversely affect mobile players' revenues.

Public Regulation

The idea behind mobile *advertising* is very similar to e-mail on the wired Internet, but with one big difference: it is "opt-in." This function is essential to give users total control over what they receive, because consumers' demand for highly personalized messages has to be reconciled with their desire for privacy (Sadeh, 2002). The Mobile Marketing Association (MMA) has attempted to establish industry guidelines for mobile marketers, as follows: (1) MMA members should not send mobile advertising without confirmed opt-in, and (2) such opt-in subscriber permission is not transferable to third parties without explicit permission from the subscriber (Petty, 2003).

Lifestyle and Habits

In general, European consumers habitually commute by car, and this provides fewer incentives to access the mobile Internet (Baldi & Thaung,

2002). In addition, a systematic "word-of-mouth" helped the rapid diffusion of i-mode in Japan, especially given the "normative beliefs attributed to significant others (friends, colleagues, or family members) with respect to adopting or continuing to use the technology" (Barnes & Huff, 2003). This may partially explain a high subscription rate (almost 75%) to e-mail newsletters among i-mode users, and this makes acceptance of mobile advertising much easier. However, this factor is unlikely to be present in many European countries, which are characterized as more individualist than Asian countries.

On the basis of the preceding discussions, the following hypotheses were formulated to test the principal thesis of the research:

- **H1:** MNCs' intention to adopt SMS-based advertising is directly and positively associated with branding technique.
- **H2:** MNCs' intention to adopt SMS-based advertising is directly and positively associated with facilitating conditions.
- **H3:** MNCs' intention to adopt SMS-based advertising is directly and positively associated with location-based services.
- **H4:** MNCs' intention to adopt SMS-based advertising is directly and negatively associated with connection costs.
- **H5:** MNCs' intention to adopt SMS-based advertising is directly and negatively associated with public regulation.
- **H6:** MNCs' intention to adopt SMS-based advertising is directly and negatively associated with lifestyle and habits.

METHODOLOGY

Questionnaire Items and Measures

A structured questionnaire was prepared, drawing on prior literature. A majority of the items were originally developed for this study, because of the scarcity of empirical research on mobile advertising. Each item was measured on a Likert-type five-point scale. A five-point scale was preferred to a seven-point scale, because telephone interviews were used, rather than a mail or other form of paper-and-pencil survey. This method was considered more appropriate because mobile advertising is still in its infancy, and company executives may not be able to make fine distinctions regarding their attitudes on this topic. During the telephone interview, interviewers followed a script. However, respondents were free to ask questions whenever they encountered definitional problems.

Table 1. Regression analysis and hypotheses testing

Hypotheses	Independent Variables		Standardized β		Results
H1	Branding technique		.553	**	Supported
H2	Facilitating conditions		.325	**	Supported
H3	Location-based services		.023		Rejected
H4	Connection costs		-.397	**	Supported
H5	Public regulation		.125		Rejected
H6	Lifestyle and habits		-.115		Rejected
		R^2	.598	**	
		ΔR^2	.013		
		ΔF	1.509		

Multinational Corporations

With regard to Japanese firms, the selection was based on the *Multinational Companies Database*. The database was created by the Research Institute for Economics and Business Administration at Kobe University (2003) and includes Japanese companies listed in the first section of the Tokyo Stock Exchange with foreign direct investment in more than five countries (Kobe University, 2003). American firms were chosen from The Forbes 500 (*Forbes*, 2003a). Finally, European firms were singled out from The Forbes International 500 (*Forbes*, 2003b), because this list indicates the nationality of each firm. Regardless of nationality, however, companies associated with aerospace and defense, food and drug retail chains, forestry and fishery, general public utilities, health care providers, heavy machines, industrial goods, local banking and insurance, metals and mining, and oil and gas extraction were excluded. Next, firms operating in Spain were identified. As a result, 43 Japanese, 47 American, and 31 European firms' Spanish subsidiaries were identified.

Telephone Interview

Telephone interviewing was considered appropriate because of the novelty of the research subject. It was expected that interviewers would be able to clarify doubts or answer any questions that interviewees might have regarding mobile communications. To this end, four bilingual assistants were employed (two Spanish and two Japanese, all fluent in English). During the second and third weeks of February 2004, intensive training was provided so that the assistants could gain sufficient skills and knowledge to conduct the telephone interview. The actual interviewing was carried out during March 2004, under the supervision of the researcher. It was established that when the target executives were absent or unavailable for interview, assistants had to ask: (1) for an appointment for the next phone call, or (2) about the availability of the person next in seniority in the marketing department to the target executive. As a result, a total of 53 interviews was conducted, with 27, 16, and 10 respondents from Japanese, American, and European firms, respectively. The response rate was 43.8%.

FINDINGS

First, an exploratory factor analysis with Equamax rotation with Kaiser Normalization was carried out. The rotation, converged in 12 iterations, produced a clear-cut six-factor solution with a cut-off value of .50. Only factors with eigenvalue greater than 1 were retained. It should be noted that the proposed construct "connection costs" was merged into a mixed construct "security and costs." However, because of the exploratory nature of the study, it was considered acceptable to use this six-factor solution for the subsequent analysis. The extracted factors explain 68.6% of the total variance, and the level of loading is consistently high across the six factors. Factor scores were retained as variables with the Anderson-Rubin method to minimize the level of multicollinearity, for the use of regression analysis. The reliability was calculated with Chronbach's alpha for each construct. The scores range from .60 to .85, exceeding the cut-off point of .60 suggested by Hair, Anderson, Tatham, and Black (1998). Next, the hypotheses were tested by performing regression analysis with a step-wise method. Each of six independent variables (i.e., factor scores) was regressed on the dependent variable, "MNCs' intention to use mobile advertising," in order of their expected contributions. The results of regression analysis are shown in Table 1.

DISCUSSION

This study aims to identify MNCs' principal perceptions of SMS-based push-type mobile

advertising and their intention to use it. On the basis of the data obtained from 53 MNCs, our principal propositions were tested by multiple regression analysis. The results were mixed: only half of the six hypotheses gained empirical support. The regression analysis identified branding technique, facilitating conditions, and connection costs as the three primary predictors influencing MNCs' intention to use mobile advertising. The contribution of branding technique in particular is substantial, indicating that MNCs are likely to perceive mobile advertising as an effective branding tool to increase brand awareness and image. Also, technological infrastructure and the availability of sophisticated mobile handsets are prerequisites for mobile marketing. As expected, unfavorable mobile Internet pricing negatively affects the MNCs' intention to use mobile advertising. On the other hand, the contributions of location-based services, public regulation, and lifestyle and habits are not only statistically insignificant, but also trivial in terms of the coefficient magnitude. One reason why location-based services were not identified as a significant factor is that the GPS system is not as widespread in Europe as it is in Japan. In addition, many Scandinavian firms, leaders of sophisticated mobile Internet service practitioners, were not included in the study. Admitting the danger of simple generalization, the findings of this study may imply that MNCs are concerned to a lesser extent with regulatory and cultural impediments to adopting mobile advertising.

REFERENCES

Baldi, S., & Thaung, H. P. P. (2002). The entertaining way to m-commerce: Japan's approach to the mobile Internet—A model for Europe? *Electronic Markets, 12*(1), 6-13.

Barnes, S. J. (2003). Wireless digital advertising: Nature and implications. *International Journal of Advertising, 21,* 399-420.

Barnes, S. J., & Huff, S. L. (2003). Rising sun: i-mode and the wireless Internet. *Communications of the ACM, 46*(11), 79-84.

Barwise, P., & Strong, C. (2002). Permission-based mobile marketing. *Journal of Interactive Marketing, 16*(1), 14-24.

Dano, M. (2002). Coke, Toyota, McDonald's test mobile advertising. *RCR Wireless News, 21*(46), 8.

Forbes. (2003a). *The Forbes 500s.* Retrieved November 3, 2003, from http://www.forbes.com/2003/03/26/500sland.html

Forbes. (2003b). *The Forbes International 500.* Retrieved November 15, 2003, from http://www.forbes.com/2003/07/07/internationaland.html

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis.* Upper Saddle River, NJ: Prentice Hall.

Kobe University. (2003). *Multinational Companies Database*. Available by permission of Research Institute for Economics and Business Administration of Kobe University. Retrieved January 11, 2004, from http://www.rieb.kobe-u.ac.jp/liaison/cdal/takokuseki/dbenterprises.html

Lu, F., Yu, C. S., Liu, C., & Yao, F. E. (2003). Technology acceptance model for wireless Internet. *Internet Research, 13*(3), 206-222.

Mylonopoulos, N. A., & Doukidis, G. I. (2003). Introduction to the special issue: Mobile business: Technological pluralism, social assimilation, and growth. *International Journal of Electronic Commerce, 8*(1), 5-22.

Petty, R. D. (2003). Wireless advertising messaging: Legal analysis directly and public policy issues. *Journal of Public Policy and Marketing, 22*(1), 71-82.

Robinson, J. P., Shaver, P. R., & Wrightsman, L.S. (1991). Criteria for scale selection and evaluation. In J. P Robinson, P. R. Shaver, & L. S. Wrights-

man (Eds.), *Measures of personality and social psychological attitudes* (pp. 1-16). San Diego: Academic Press.

Sadeh, N. (2002). *M-commerce: Technologies, services, and business models.* New York: John Wiley & Sons.

Scuka, D. (2003). *How Europe really differs from Japan*. Retrieved February 11, 2004, from http://www.mobiliser.org/article?id=68

KEY TERMS

Barcode Mobile Coupon: Mobile barcoding can be used in the form of a picture SMS which is delivered to a mobile phone. Recipients save the image, arrive at the destination, and present their barcode SMS to be scanned.

i-Mode: A broad range of Internet services for a monthly fee of approximately three euro, including e-mail, transaction services (e.g., banking, trading, shopping, ticket reservations, etc.), infotainment services (e.g., news, weather, sports, games, music download, karaoke, etc.), and directory services (e.g., telephone directory, restaurant guide, city information, etc.), which offers more than 3,000 official sites accessible through the i-mode menu.

Push Messaging Service: Various forms of messaging services are generally offered in mobile Internet. For example, SMS and WAP Push messaging generally allow users to send 100-160 characters, while mobile e-mail in Japanese i-mode allows up to 1,000 characters.

SPAM: Unsolicited or undesired bulk electronic messages. Because of the development of anti-SPAM programs, they are often deleted without being opened.

SPIM: A variation of SPAM through instant messaging systems.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 885-888, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 6.4
V-Card:
Mobile Multimedia for Mobile Marketing

Holger Nösekabel
University of Passau, Germany

Wolfgang Röckelein
EMPRISE Consulting Düseldorf, Germany

ABSTRACT

This chapter presents the use of mobile multimedia for marketing purposes. Using V-Card, a service to create personalized multimedia messages, as an example, the advantages of sponsored messaging are illustrated. Benefits of employing multimedia technologies, such as mobile video streaming, include an increased perceived value of the message and the opportunity for companies to enhance their product presentation. Topics of discussion include related projects, as marketing campaigns utilizing SMS and MMS are becoming more popular, the technical infrastructure of the V-card system, and an outline of social and legal issues emerging from mobile marketing. As V-card has already been evaluated in a field test, these results can be implemented to outline future research and development aspects for this area.

INTRODUCTION

The chapter presents the use of mobile multimedia for marketing purposes, specifically focusing on the implementation of streaming technologies. Using V-card, a service for creating personalized multimedia messages, as an example, the advantages of sponsored messaging are illustrated. Topics of discussion include related projects, as marketing campaigns utilizing SMS and MMS are becoming more popular, the technical infrastructure of the V-card system, and an outline of social and legal issues emerging from mobile marketing. As V-card has already been evaluated in a field test, these results can be implemented to outline future research and development aspects for this area.

Euphoria regarding the introduction of the universal mobile telephony system (UMTS) has evaporated. Expectations about new UMTS services are rather low. A “killer application” for 3rd generation networks is not in sight. Users are primarily interested in entertainment and news, but only few of them are actually willing to spend money on mobile services beyond telephony. However, for marketing campaigns the ability to address specific users with multimedia content holds an interesting perspective.

Advertisement-driven sponsoring models will spread in this area, as they provide benefits

to consumers, network providers, and sponsors. Sponsoring encompasses not only a distribution of pre-produced multimedia content (e.g., by offering wallpapers), Java games, or ringtones based on a product, but also mobile multimedia services.

Mobile multimedia poses several problems for the user. First, how can multimedia content of high quality be produced with a mobile device. Cameras in mobile telephones are getting better with each device generation; still the achievable resolutions and framerates are behind the capabilities of current digital cameras. Second, how can multimedia content be stored on or transmitted from a mobile device. Multimedia data, sophisticated compression algorithms notwithstanding, is still large, especially when compared to simple text messages. External media, such as memory cards or the Universal Media Disk (UMD), can be used to a certain degree to archive and distribute data. They do not provide a solution for spreading this data via a wireless network to other users. Third, editing multimedia content on mobile devices is nearly impossible. Tools exist for basic image manipulation, but again their functionality is reduced and handling is complex.

Kindberg, Spasojevic, Fleck, and Sellen (2005) found in their study that camera phones are primarily used to capture still images for sentimental, personal reasons. These pictures are intended to be shared, and sharing mostly takes place in face-to-face meetings. Sending a picture via e-mail or MMS to a remote phone occurred only in 20% of all taken pictures. Therefore, one possible conclusion is that users have a desire to share personal moments with others, but current cost structures prohibit remote sharing and foster transmission of pictures via Bluetooth or infrared.

V-card sets out to address these problems by providing a message-hub for sublimated multimedia messaging. With V-card, users can create personalized, high-quality multimedia messages (MMS) and send those to their friends. Memory constraints are evaded by implementing streaming audio and video where applicable. V-cards can consist of pictures, audio, video, and MIDlets (Java 2 Micro-Edition applications). Experience with mobile greetingcards show that users are interested in high-quality content and tend to forward them to friends and relatives. This viral messaging effect increases utilisation of the V-card system and spreads the information of the sponsor. Haig (2002, p. 35) lists advice for successful viral marketing campaigns, among them:

- Create of a consumer-to-consumer environment
- Surprise the consumers
- Encourage interactivity

A V-card message is sponsored, but originates from one user and is sent to another user. Sponsoring companies therefore are actually not included in the communication process, as they are neither a sender nor a receiver. V-card is thus a true consumer-to-consumer environment. It also can be expected for the near future that high quality content contains an element of surprise, as it exceeds the current state of the art of text messaging. Interactivity is fostered by interesting content, which is passed on, but also by interactive elements like MIDlet games.

Additionally, Lippert (2002) presents a "4P strategy" for mobile advertising, listing four characteristics a marketing campaign must have:

- Permitted
- Polite
- Profiled
- Paid

"Permitted" means a user must agree to receive marketing messages. With V-card, the originator of the MMS is not a marketing company but another user, therefore the communication itself is emphasized, not the marketing proposition. Legal aspects regarding permissions are discussed detailed below. Marketing messages should also be "polite," and not intrusive. Again, the enhanced

multimedia communication between the sender and the receiver is in the foreground, not the message from the sponsor.

"Profiling" marketing tools enables targeted marketing and avoids losses due to non-selective advertising. Even if V-card itself is unable to match a sponsor to users, since users do not create a profile with detailed personal data, profiling is achieved by a selection process of the sender. As messages can be enhanced by V-card with media related to a specific sponsor, by choosing the desired theme the sender tailors a message to the interests of himself and the receiver. Usually, marketing messages should provide a target group with incentives to use the advertised service; the recipients need to get "paid." With V-card, sponsors "pay" both users by reducing the costs of a message and by providing high quality multimedia content.

V-CARD ARCHITECTURE

V-Card Core Architecture

Figure 1 shows the V-card core architecture and illustrates the workflow. First, the user with a mobile device requests a personalised application via the SMSC or Multimedia Messaging Service Centre (MMSC), which are part of the mobile network infrastructure. The message is passed on to the V-card core, where the connector decides which application has been called.

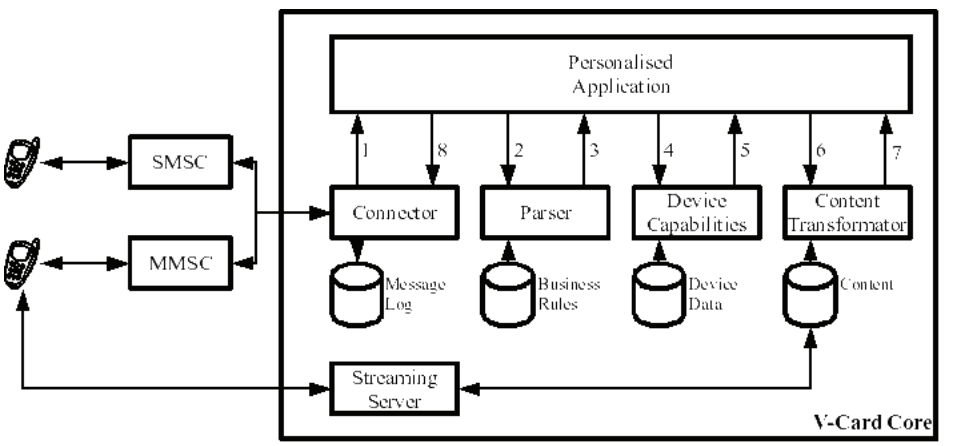
After the request is passed on to the appropriate application (1), it is logged in the message log. A parser receives the message (2), extracts the relevant data for customisation, and returns this data (3)–this could include the receiver's phone number, the name of the sender or a message. Then, the capabilities of the receiving phone are queried from a database which holds all relevant data (4+5) like display size, number of colours, supported video and audio codecs.

Finally, the application transmits all the data gathered to the content transformator. Here, the pre-produced content is tailored with the input delivered by the user according to the capabilities of the device (6+7). The result is then sent via the connector (8) to the receiving user. Since the personalised applications and the data are separated, new applications can be easily created.

V-Card Streaming Technology

Since video content can not be stored directly on a mobile device due to memory limitations, a streaming server supplies video data to the device where the video is played, but not stored with

Figure 1. V-Card core architecture



the exception of buffered data, which is stored temporarily to compensate for varying network throughput. Streaming video and audio to mobile devices can be utilized for various services (e.g., for mobile education) (Lehner, Nösekabel, & Schäfer 2003). In the case of V-card, the MMS contains a link to adapted content stored on the content server. This link can be activated by the user and is valid for a certain amount of time. After the link has expired, the content is removed from the content server to conserve memory.

Currently, there are two streaming server solutions available for mobile devices. RealNetworks offers the HELIX server based on the ReadMedia format. RealPlayer, a client capable of playing back this format, is available for Symbian OS, Palm OS 5, and PocketPC for PDAs. Additionally, it is available on selected handsets, including the Nokia 9200 Series Communicators and Nokia Series 60 phones, including the Nokia 7650 and 3650. The other solution is using a standardized 3GPP-stream based on the MPEG4 format, which can be delivered using Apples Darwin server.

An advantage of implementing streaming technology for mobile multimedia is the fact that only a portion of the entire data needs to be transmitted to the client, and content can be played during transmission. Data buffers ensure smooth playback even during short network interruptions or fluctuations in the available bandwidth. As video and audio are time critical, it is necessary that the technologies used are able to handle loss of data segments, which do not arrive in time (latency) or which are not transmitted at all (network failure). GPRS and HSCSD connections allow about 10 frames per second at a resolution of 176 by 144 pixel (quarter common intermediate format QCIF resolution) when about 10 KBit per second are used for audio. Third generation networks provide a higher bandwidth, leading to a better quality and more stable connectivity.

A drawback of streaming is the bandwidth requirement. For one, the bandwidth should be constant; otherwise the buffered data is unable to compensate irregularities. Next, the available bandwidth directly influences the quality that can be achieved — the higher the bandwidth, the better the quality. Third, a transfer of mobile data can be expensive. A comparison of German network providers in 2003 showed that 10 minutes of data transfer at the speed of 28 KBit per second (a total amount of 19 Megabyte) resulted in costs ranging from 1 Euro (time-based HSCSD tariff) up to 60 Euro (packet-based GPRS by call tariff).

V-Card Examples

Figures 2 demonstrates a picture taken with the camera of a mobile device, rendered into a video clip by the V-card core. Figure 3 combines pictures and text from the user with video and audio content from the V-card hub. Figure 4 shows how simple text messages can be upgraded when a picture and an audio clip are added to create a multimedia message.

Since sponsoring models either influence the choice of media used to enhance a message, or can be included as short trailers before and after the actual message, users and sponsors can choose from a wide variety of options best suited for their needs.

LEGAL ASPECTS

It should be noted that the following discussion focuses on an implementation in Germany and today (2005Q1)–although several EU guidelines are applicable in this area there are differences in their national law implementations and new German and EU laws in relevant areas are pending.

Legal aspects affect V-card in several areas: consumer information laws and rights of withdrawal, protection of minors, spam law, liability, and privacy. A basic topic to those subjects is the classification of V-card among "Broadcast Service" ("Mediendienst"), "Tele Service" ("Teledienst"), and "Tele Communication Service"

Figure 2. V-Card with picture in video

Figure 3. V-Card with picture and text in video

Figure 4. V-Card with text in picture and audio

("Telekommunikationsdienst"). According to § 2 Abs. 2 Nr. 1 and § 2 Abs. 4 Nr. 3 Teledienstegesetz (TDG) V-card is not a "Broadcast Service" and based on a functional distinction (see e.g., Moritz/Scheffelt in Hoeren/Sieber, 4, II, Rn. 10) V-card is presumed to be a "Tele Service."

Consumer information laws demand that the customer is informed on the identity of the vendor according to Art. 5 EU Council Decision 2000/31/EC, to § 6 TDG and to § 312c Bürgerliches Gesetzbuch (BGB) (e.g., on certain rights he has with regard to withdrawal). The fact that V-card might be free of charge for the consumer does not change applicable customer protection laws as there is still a (one-sided) contract between the costumer and the provider (see e.g., Bundesrat, 1996, p. 23). Some of these information duties have to be fulfilled before contract and some after. The post-contract information could be included in the result MMS and the general provider information and the pre-contract information could be included in the initial advertisements and/or a referenced WWW- or WAP-site. Art. 6 EU Council Decision 2000/31/EC and § 7 TDG demand a distinction between information and adverts on Web sites and can be applicable, too. A solution could be to clearly communicate the fact that the V-card message contains advert (e.g., in the subject) (analogue to Art. 7(1) EU Council Decision 2000/31/EC, although this article is not relevant in Germany). The consumer might have a withdrawal right based on § 312d (1) BGB on which he has to be informed although the exceptions from § 312c (2) 2nd sentence BGB or § 312d (3) 2 BGB could be applicable. With newest legislation the consumer has to be informed on the status of the withdrawal rights according to § 1 (1) 10 BGB-Informationspflichtenverordnung (BGB-InfoV), whether he has withdrawal rights or not.

§ 6 Abs. 5 Jugendmedienschutzstaatsvertrag (JMStV) bans advertisements for alcohol or tobacco which addresses minors, § 22 Gesetz über den Verkehr mit Lebensmitteln, Tabakerzeugnissen, kosmetischen Mitteln und sonstigen

Bedarfsgegenständen (LMBG) bans certain kinds of advertisements for tobacco, Art. 3(2) EU Council Decision 2003/33/EC (still pending German national law implementation) bans advertisements for tobacco in Tele Services. Therefore a sponsor with alcohol or tobacco products will be difficult for V-card. Sponsors with erotic or extreme political content will also be difficult according to § 4, 5 and 6(3) JMStV. § 12(2) 3rd sentence Jugendschutzgesetz (JuSchG) demands a labelling with age rating for content in Tele Services in case it is identical to content available on physical media. Since V-card content will most of the time special-made and therefore not available on physical media, this is not relevant.

The e-mail spam flood has led to several EU and national laws and court decisions trying to limit spam. Some of these laws might be applicable for mobile messaging and V-card, too. In Germany a new § 7 in the Gesetz gegen den unlauteren Wettbewerb (UWG) has been introduced. The question in this area is whether it can be assumed that the sent MMS is ok with the recipient (i.e., if an implied consent can be assumed). Besides the new § 7 UWG if the implied consent cannot be assumed a competitor or a consumer rights protection group could demand to stop the service because of a "Eingriff in den eingerichteten und ausgeübte Gewerbebetrieb" resp. a "Eingriff in das Allgemeine Persönlichkeitsrecht des Empfängers" according to §§ 1004 resp. 823 BGB.

Both the new § 7 UWG and previous court decisions focus on the term of an unacceptable annoyance or damnification which goes along with the reception of the MMS. The highest German civil court has ruled in a comparable case of advert sponsored telephone calls (BGH reference I ZR 227/99) that such an implied consent can be assumed under certain conditions e.g. that the communication starts with a private part (and not with the advertisement) and that the advertisement is not a direct sales pitch putting psychological pressure on the recipient (see e.g., Lange 2002, p. 786). Therefore if a V-card message consists of a private part together with attractive and entertaining content and a logo of the sponsor the implied consent can be assumed. The bigger the advertisement content part is the likelier it is that the level of a minor annoyance is crossed and the message is not allowed according to § 7 UWG (see e.g., Harte-Bavendamm & Henning-Bodewig, 2004, § 7, Rn. 171).

If users use the V-card service to send unwelcome messages to recipients V-card could be held responsible as an alternative to the user from whom the message originated. A Munich court (OLG München reference 8 U 4223/03) ruled in this direction in a similar case of an e-mail news letter service however focusing on the fact that the service allowed the user to stay anonymously. This is not the case with the mobile telephone numbers used in V-card, which are required to be associated with an identified person in Germany. In addition to this the highest German court has in some recent decisions (BGH I ZR 304/01, p. 19 and I ZR 317/01, p. 10) narrowed the possibilities for a liability as an alternative by limiting the reasonable examination duties.

Manual filtering by the V-card service is a violation of communication secrecy and therefore not allowed (see e.g., Katernberg, 2003). Automatic filtering must not result in message suppression since this would be illegal according to German martial law § 206 (2) 2 Strafgesetzbuch.

The obligation to observe confidentiality has in Germany the primary rule that data recording is not allowed unless explicitly approved (§ 4 Bundesdatenschutzgesetz). Log files would therefore not be allowed with an exception for billing according to § 6 Gesetz über den Datenschutz bei Teleddiensten (TDDSG). These billing logs must not be handed over to third parties likely also including the sponsor.

As a conclusion, it can be noted that an innovative service like V-card faces numerous legal problems. During the project, however, it became clear that all these requirements can be met by an appropriate construction of the service.

EVALUATION OF V-CARD

Since V-card also has the ability to transmit personalised J2ME applications via MMS (see Figure 5 for an example), it surpasses the capabilities of pure MMS messages creating added value for the user, which normally do not have the possibility to create or modify Java programs. One example is a sliding puzzle where, after solving the puzzle, a user may use the digital camera of the mobile device to change the picture of the puzzle. After the modification, the new puzzle can then be send via V-card to other receivers.

Still, as previously mentioned, V-card requires a MMS client. It can therefore be regarded as an enhancement or improvement for MMS communication and is as such a competitor to the "normal" MMS. Hence, an evaluation framework should be usable to measure the acceptance of both "normal" MMS messaging and "enhanced" V-card messaging, creating results that can be compared with each other to determine the actual effect of the added value hoped to be achieved with V-card. While extensive research exists regarding PC-based software, mobile applications currently lack comprehensive methods for creating such evaluations. Therefore, one possible method was developed and applied in a fieldtest to evaluate V-card (Lehner, Sperger, & Nösekabel, 2004).

At the end of the project on June 3, 2004, a group of 27 students evaluated the developed V-card applications in a fieldtest. Even though the composition and size of the group does not permit to denote the results as representative, tendencies can be identified. The statistical overall probability of an error is 30%, as previously mentioned.

The questionnaire was implemented as an instrument to measure results. To verify the quality and reliability of the instrument, three values were calculated based on the statistical data. The questionnaire achieved a Cronbach alpha value of 0.89—values between 0.8 and 1.0 are regarded as acceptable (Cronbach, 1951). The split-half correlation, which measures the internal consistency

Figure 5. V-card with MIDlet puzzle application

of the items in the questionnaire, was calculated to be 0.77 with a theoretical maximum of 1.0. Using the Spearman-Brown formula to assess the reliability of the instrument, a value of 0.87 was achieved. Again, the theoretical maximum is 1.0. Therefore, the questionnaire can be regarded to be statistically valid and reliable.

One result of the fieldtest was that none of the students encountered difficulties in using any of the V-card applications, even though the usability of the mobile phone used in the fieldtest was regarded as less than optimal.

Overall 66% of the students thought that V-card was easy to use, 21% were undecided. It is very likely that the sample group leaned towards a negative or at least neutral rating as the usability of the end device was often criticised. This factor can not be compensated by the programmers of the mobile application. Another indicator for this rationale is the comparison with the results for the MMS client. Here, 75% of the group agreed to this statement, which is an increase of 9%. The similarity of the results suggests that also the rating for the usability of the MMS client was tainted by the usability of the device.

No uniform opinion exists regarding sponsored messages by incorporating advertising. Forty-two

percent of the students would accept advertisements if that would lower the price of a message. Thirty-seven percent rejected such a method. The acceptable price for a V-card message was slightly lower compared to that of a non-sublimated MMS, which on the other hand did not contain content from a sponsor.

An important aspect for the acceptance of mobile marketing is the protection of privacy. In this area the students were rather critical. Sixty-three percent would reject to submit personal data to the provider of V-card. Since this information was not necessary to use V-card, only 17% of the sample group had privacy concerns while using V-card.

The mobile marketing component was perceived by all participants and was also accepted as a mean to reduce costs. This reduction should benefit the user, therefore a larger portion of the sample group rejected for V-card the idea for increased cost incurred by a longer or more intensive usage (88% rejected this for V-card, 67% for MMS).

As already addressed, the pre-produced content of V-card helped 50% of the users to achieve the desired results. The portion rejecting this statement for V-card was 25%, which is higher than the 8% who rejected this statement for MMS. This leads to the conclusion that if the pre-produced content is appropriate in topic and design for the intended message, it contributes to the desired message. However, it is not possible to add own content if the pre-produced content and the intention of the sender deviate. The user is therefore limited to the offered media of the service provider.

Overall, the ratings for V-card by the students were positive. Marketing messages, which were integrated into the communication during the fieldtest, were not deemed objectionable. The usability of V-card was also rated high. Main points that could be addressed during the actual implementation in the mobile market should include privacy and cost issues.

CONCLUSION

The new messaging service MMS has high potential and is being widely adopted today, although prices and availability are far from optimal. Mostly young people tend to use the fashionable messages which allow much richer content to be sent instantly to a friend's phone. This young user group is especially vulnerable to debts due to their mobile phones though, or they have prepaid subscriptions letting them only send a very limited number of messages. By incorporating a sponsor model in V-card, this user group will be able to send a larger number of messages with no additional cost and thereby offering advertising firms a possibility to market their services and goods. For those users that are not as price sensitive, the large amount of professional media and the ease of the message-composition will be an incentive to use the service. The added value of the service should be a good enough reason to accept a small amount of marketing in the messages. Since V-card offers the sender and receiver an added value, the marketing message will be more acceptable than other forms of advertising where only the sender benefits from the advertisement.

Another advantage of V-card is the fact that the system takes care of the administration and storing of professional media and the complicated formatting of whole messages, thus taking these burdens from the subscriber. At the same time, V-card offers marketers a new way to reach potential customers and to keep in dialogue with existing ones. The ease of sending such rich content messages with a professional touch to a low price or even no cost at all will convince subscribers and help push 3G networks.

Overall, it can be expected that marketing campaigns will make further use of mobile multimedia streaming, aided by available data rates and the increasing computing power of mobile devices. Continuous media (video and audio), either delivered in real-time or on demand, will possibly become the next entertainment paradigm for a mobile community.

REFERENCES

Bundesrat. (1996). *Bundesrats-Drucksache* 966/96. Köln: Bundesanzeiger Verlagsgesellschaft mbH.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Haig, H. (2002). *Mobile marketing—The message revolution*. London: Kogan Page.

Harte-Bavendamm, H., & Henning-Bodewig, F. (2004). *UWG Kommentar.* München: Beck.

Hoeren, T., & Sieber, U. (2005). *Handbuch Multimedia-Recht.* München: Beck.

Katernberg, J. (2003). *Viren-Schutz/Spam-Schutz.* Retrieved from http://www.uni-muenster.de/ZIV/Hinweise/Rechtsgrundlage VirenSpamSchutz.html

Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005). The ubiquitous camera: An in-depth study of camera phone use. *IEEE Pervasive Computing, 4*(2), 42-50.

Lehner, F., Nösekabel, H., & Schäfer, K. J. (2003). *Szenarien und Beispiele für Mobiles Lernen*. Regensburg: Research Paper of the Chair of Business Computing III Nr. 67.

Lehner, F., Sperger, E. M., & Nösekabel, H. (2004). Evaluation framework for a mobile marketing application in 3rd generation networks. In K. Pousttchi, & K. Turowski (Eds.), *Mobile Economy—Transaktionen, Prozesse, Anwendungen und Dienste* (pp.114-126). Bonn: Köllen Druck+Verlag.

Lange, W. (2002). Werbefinanzierte Kommunikationsdienstleistungen. *Wettbewerb in Recht und Praxis, 48*(8), 786-788.

Lippert, I. (2002). Mobile marketing. In W. Gora, & S. Röttger-Gerigk (Eds.), *Handbuch Mobile-Commerce* (pp.135-146). Berlin: Springer.

KEY TERMS

MMS: Multimedia message service: Extension to SMS. A MMS may include multimedia content (videos, pictures, audio) and formatting instructions for the text.

Multimedia: Combination of multiple media, which can be continuous (e.g., video, audio) or discontinuous (e.g., text, pictures).

SMS (Short Message Service): text messages that are sent to a mobile device. A SMS may contain up to 160 characters with 7-bit length, longer messages can be split into multiple SMS.

Streaming: Continuous transmission of data primarily used to distribute large quantities of multimedia content.

UMTS (Universal Mobile Telecommunications System): 3rd generation network, providing higher bandwidth than earlier digital networks (e.g., GSM, GPRS, or HSCSD).

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 430-439, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.5
Mobile Multimedia for Commerce

P.M. Melliar-Smith
University of California, Santa Barbara, USA

L.E. Moser
University of California, Santa Barbara, USA

INTRODUCTION

The ready availability of mobile multimedia computing and communication devices is driving their use in commercial transactions. Mobile devices are lightweight and wireless so users can carry them and move about freely. Such devices include cell phones, PDAs and PCs equipped with cellular modems.

In the history of man, mobile commerce was the conventional form of commerce but during the twentieth century, it was superseded by fixed locations as a result of non-mobile infrastructure (stores and offices) and the ability of customers to travel. With modern mobile infrastructure, commerce can be conducted wherever the customer is located, and the sales activity can occur wherever and whenever it is convenient for the customer.

BACKGROUND

Mobile computing and communication devices, based on cellular communication, are a relatively recent innovation. Multimedia computing and communication, including video, audio, and text, are available for mobile devices but are limited by small screens, low bandwidth, and high transmission costs. These limitations distinguish mobile multimedia computing and communication from desktop multimedia computing and communication over the Internet, including WiFi, and dictate a somewhat different approach.

Mobile commercial processes are still largely experimental and are not yet well established in practice. Some researchers (Varshney, 2000) have projected that the use of mobile devices in consumer-to-business transactions will increase as much as 40%. Cautious consumers, inadequate mobile devices, security concerns, and undevel-

oped business models and procedures currently limit the use of mobile multimedia devices for commercial transactions.

Because mobile multimedia commerce using mobile devices is a new and developing field, there is relatively little available information, and that information is scattered. Early discussions of mobile commerce can be found in Senn (2000) and Varshney (2000). The i-mode service (Kinoshita, 2002; Lane, 2002) for mobile commerce has achieved some commercial success, within the limitations of existing devices and protocols.

LIMITATIONS OF THE MOBILE DEVICE

Cellular communication is wireless communication between mobile devices (e.g., cell phones, PDAs, and PCs) and fixed base stations. A base station serves mobile devices within a relatively small area of a few square miles (called a cell). The base stations are interconnected by fixed telecommunication infrastructure that provides connection with other telecommunication systems. When a mobile device passes from the cell of one base station to that of another, the first base station hands off communication with the device to the other, without disrupting communication.

Mobile devices are inherently more limited than fixed devices, but these limitations, appropriately recognized and accommodated, do not preclude their use in commerce (Buranatrived, 2002; Lee & Benbasat, 2003). Mobile devices have restricted display, input, print, and communication capabilities. The impact of these limitations depends on the user. A professional mobile sales representative needs better display, input, and print capabilities than many other kinds of users.

Mobile devices, such as cell phones and PDAs, have very small displays (less than 15 cm) that are likely to remain small, a limitation imposed by the need to insert the device into a pocket or purse, or to carry the device on a belt, and also by battery consumption. Such displays are inadequate for viewing detailed textual or graphical material. In an environment that is saturated with television, video, animated Web pages, and so forth, impressive multimedia sales presentations are even more important. Therefore, a mobile sales representative most likely will carry a notebook computer with a high-resolution display of 30 cm to 50 cm, and might even carry a projection display, which imposes little limitation on the material to be displayed.

The input capabilities of current mobile devices, such as cell phones and PDAs, are currently primitive and difficult to use for commercial activities. When natural language voice input is improved, the input of more complex requests, responses, and textual material will be possible. Substantial advances in speech recognition and natural language processing are necessary, and substantial increases in processing power and battery capacity are required before this promise can be realized.

Mobile devices are unlikely to provide printed output, but a mobile sales representative will likely carry a portable printer with which to create documents for the customer. Alternatively, such documents might be transferred directly between the mobile sales representative's device and the customer's device, using a cellular, infrared, bluetooth, or other wireless connection, without a physical paper record.

Storage capacity is not really a limitation for mobile commerce; hard disk capacities of many Gbytes are available for mobile devices. Similarly, the bandwidth of cellular communication links is sufficient for commercial interactions; however, the cost of transmitting detailed graphics over a cellular link is relatively high. Therefore, a mobile commercial sales representative will likely carry, on hard disk or CD, presentations and catalogs that contain detailed graphics or video, so that they do not need to be downloaded over an expensive wireless connection.

Typical mobile devices operate with low bandwidth, too low to allow effective display of video or Web pages. Remarkable efforts have been made with i-mode services (Kinoshita, 2002; Lane, 2002) to achieve effective mobile commerce, despite bandwidth limitations. The 3G networks currently being deployed provide sufficient bandwidth for display of video and Web pages. However, the high cost of cellular communication remains a significant limitation on activities that require large amounts of information to be transmitted. Mobile commercial activities need to operate with minimal or intermittent connections and with activities conducted while disconnected.

Currently, battery power and life are also significant limitations on mobile multimedia devices, restricting the availability of processing, display, and communication. However, small, light, mobile, alcohol-based fuel cells are in prototype and demonstration. When substantial demand develops for more powerful mobile multimedia devices, more powerful batteries will become available.

NEEDS OF USERS

It is important to distinguish between the needs of sellers and buyers and, in particular, the needs of:

- Professional mobile sellers;
- Professional mobile buyers;
- Convenience purchasers.

The popular concept of mobile commerce focuses on the buyer, but buyers are motivated by convenience, and attractive, effective capabilities are required to achieve significant adoption by buyers. In contrast, sellers are motivated by need, and they are more likely to be early adopters of novel technology.

Needs of Mobile Sellers

Professional mobile sellers include insurance agents, contractors, and other sales people who make presentations on the customers' premises. In the Internet era, with customers who do not need to visit a seller to make a purchase, sellers no longer need to wait for customers but need to become mobile to find customers wherever they can be found. Mobile sellers require support for contact information, appointments, scheduling, and reminders. PC-based tools provide such services, although their human interfaces are not appropriate for mobile devices. Mobile sales people might also use Customer Relationship Management (CRM) software that likely will run on a central server and will be accessed remotely by a seller using a cellular Internet connection.

The most demanding aspect of the work of a mobile sales person is the presentation to the customer. A mobile sales person lacks the large physical stock and demonstration models available at a fixed site but, instead, must depend on a computer-generated display of the product. An impressive multimedia presentation is essential for selling in an environment that is saturated with television, video, animated Web pages, and the like. Thus, a mobile sales person can be expected to carry a display device (a laptop computer or a projection display), with presentations and catalogs stored on hard disk or CD. Significant effort is required to make a computer-hosted catalog as convenient to use as a conventional paper catalog, but a large computer-hosted catalog is more convenient to carry, can be searched, can contain animations, and can be updated more easily and more frequently.

Access to a catalog or other presentation material hosted on a central server is unattractive because of the cost and time of downloading detailed graphic presentations over an expensive wireless link. However, a mobile sales person needs a cellular Internet communication with the central server to query inventory, pricing, and delivery;

to enter sales orders; to make reservations; and to schedule fulfillment of the sale. The mobile sales person also needs to generate proposals and contracts on the mobile device and print them for the customer. Many customers will accept electronic delivery of proposals and exchange of contracts; however, some customers will require paper copies. and, thus, the mobile sales person must carry a printer.

In summary, a laptop computer with a cellular modem and a portable printer, possibly augmented by a projection display for multimedia presentations, can satisfy the needs of a mobile sales person.

Needs of Mobile Buyers

The direct mobile buyer analog of the mobile sales person, a buyer who visits sellers to purchase goods, such as a buyer who visits ranchers to purchase livestock or visits artists to purchase paintings, is unlikely to develop. Such sellers have already discovered the use of the Internet to sell their products at higher prices than such a visiting buyer would offer.

Professionals who need to purchase while mobile include contractors and travelers. When using a mobile device such as a cell phone or PDA, they are likely to limit their activities to designation of items and quantities, delivery address and date, and payment information. They are unlikely to use such mobile devices to browse catalogs and select appropriate merchandise, because of the inadequate display and input capabilities of the devices and because of the cost of cellular Internet connections. It is essential to analyze carefully the model of transactions in a specific field of commerce, and the software and interactions needed to support that model (Keng, 2002).

Current cell phones and PDAs are barely adequate in their input and output capabilities for purchase of items in the field (Buranatrived, 2002; Lee, 2003). The small display size of portable mobile devices is unlikely to change soon but can be compensated to some extent by Web pages that are designed specifically for those devices. Such mobile-friendly Web pages must be designed not only to remove bandwidth-hogging multimedia and graphics and to reduce the amount of information presented, but also to accommodate a professional who needs to order items with minimum interaction.

Web pages designed originally for high-resolution desktop computers can be downgraded automatically, so that they require less transmission bandwidth. However, such automated downgrading does not address the abbreviated interaction sequences needed by a professional using a mobile device. Most professionals would prefer to make a conventional phone call to purchase goods, rather than to use an existing mobile device.

Mobile devices such as cell phones and PDAs have inadequate input capabilities for such mobile buyers, particularly when they are used in restricted settings such as a building site or a moving truck. This problem will be alleviated by natural language voice input when it becomes good enough. Until then, professional mobile buyers might prefer to select a small set of items from the catalog, download them in advance to the mobile device using a fixed infrastructure communication link, and use a retrieval and order program specifically designed for accessing the downloaded catalog items on the mobile device.

Needs of Convenience Purchasers

Convenience purchasers expect simpler human interfaces and lower costs than professional sellers or buyers (Tarasewich, 2003). For the convenience purchaser, because of the poor human interfaces of current mobile devices, a purely digital mobile commercial transaction is substantially less convenient and satisfying than visiting a store, making a conventional telephone call, or using the better display, easier interfaces, and lower costs of a PC to purchase over the Internet.

Convenience purchasers are most likely to purchase products that are simple and highly standardized, or that are needed while mobile. Nonetheless, mobile devices can facilitate commercial transactions in ways other than direct purchase. For example, a mobile device associated with its human owner can be used to authorize payments in a way that is more convenient than a credit card (Ogawara, 2002).

The mobile device is usually thought of as facilitating commercial transactions through mobility in space, but locating a customer in space is also an important capability (Bharat, 2003). However, location-aware services typically benefit the seller rather than the mobile purchaser, and somewhat resemble spam. A mobile device also can be used to facilitate the collection of information through time, particularly if the device is continuously present with and available to its user.

ENABLING TECHNOLOGY FOR MOBILE MULTIMEDIA

The Wireless Application Protocol, Wireless Markup Language, and Wireless Security Transport Layer discussed next are used in commercial mobile devices and enable the use of mobile multimedia for commerce.

Wireless Application Protocol

The Wireless Application Protocol (WAP) is a complex family of protocols (WAP Forum, 2004), for mobile cell phones, pagers, and other wireless terminals. WAP provides:

- Content adaptation, using the Wireless Markup Language (WML) discussed later, and the WMLScript language, a scripting language similar to JavaScript that is oriented toward displaying pages on small low-resolution displays.
- Reliability for display of Web pages provided by the Wireless Datagram Protocol (WDP) and the Wireless Session Protocol (WSP) to cope with wireless connections that are rather noisy and unreliable.
- Efficiency, provided by the WDP and the WSP through data and header compression to reduce the bandwidth required by the applications.
- Integration of Web pages and applications with telephony services provided by the WSP and the Wireless Application Environment (WAE), which allows the creation of applications that can be run on any mobile device that supports WAP.

Unfortunately, WAP's low resolution and low bandwidth are traded off against convenience of use. Because screens are small and input devices are primitive, selection of a service typically requires inconvenient, confusing, and time-consuming steps down a deep menu structure. Successful applications have been restricted to:

- Highly goal-driven services aimed at providing immediate answers to specific problems, such as, "My flight was canceled; make a new airline reservation for me."
- Entertainment-focused services, such as games, music, and sports, which depend on multimedia.

As mobile devices become more capable, WAP applications will become easier to use and more successful.

Wireless Markup Language (WML)

The Wireless Markup Language (WML), which is based on XML, describes Web pages for low-bandwidth mobile devices, such as cell phones. WML provides:

- Text presentation and layout – WML includes text and image support, including a variety of format and layout commands, generally simple and austere, as befits a small screen.
- Deck/card organizational metaphor – in WML, information is organized into a collection of cards and decks.
- Intercard navigation and linking – WML includes support for managing the navigation between cards and decks with reuse of cards to minimize markup code size.
- String parameterization and state management – WML decks can be parameterized using a state model.
- Cascading style sheets – these style sheets separate style attributes for WML documents from markup code, reducing the size of the markup code that is transmitted over a cellular link and that is stored in the memory of the mobile device.

WML is designed to accommodate the constraints of mobile devices, which include the small display, narrow band network connection, and limited memory and computational resources. In particular, the binary representation of WML, as an alternative to the usual textual representation, can reduce the size of WML page descriptions.

Unfortunately, effective display of pages on low-resolution screens of widely different capabilities requires WML pages that are specifically, individually, and expensively designed for each different mobile device, of which there are many. In contrast, HTML allows a single definition for a Web page, even though that page is to be viewed using many kinds of browsers and displays.

Wireless Transport Security Layer

Security is a major consideration in the design of systems that provide mobile multimedia for commerce. The Wireless Transport Security Layer (WTSL) aims to provide authentication, authorization, confidentiality, integrity, and non-repudiation (Kwok-Yan, 2003; WAPForum, 2004; Wen, 2002). Major concerns are:

- Disclosure of confidential information by interception of wireless traffic, which is addressed by strong encryption.
- Disclosure of confidential information, including location information within the wireless service provider's WAP gateway, which can be handled by providing one's own gateway, although most users might prefer to rely on the integrity of the wireless service provider.
- Generation of transactions that purport to have been originated by a different user, which can be handled by Wireless Identity Modules (WIMs). A WIM, which is similar to a smart card and can be inserted into a WAP-enabled phone, uses encryption with ultra-long keys to provide secure authentication between a client and a server and digital signatures for individual transactions. WIMs also provide protection against interception and replay of passwords.
- Theft and misuse of the mobile device, or covert Trojan horse code that can extract encryption keys, passwords, and other confidential information from the mobile device, which is handled by WIMs that can be but probably will not be removed from the mobile device for safe keeping, and that can themselves be lost or stolen.

WTSL is probably provides adequate security for most commercial mobile multimedia transactions, and is certainly more secure than the vulnerable credit card system that is used today for many commercial transactions.

CONCLUSION

Mobile multimedia will be a significant enabler of commerce in the future, as mobile devices become more capable, as multimedia provides

more friendly user interfaces and experiences for the users, and as novel business models are developed. Great care must be taken to design services for mobile multimedia commerce for the benefit of the mobile user rather than the sellers of the service. Natural language voice input and intelligent software agents will increase the convenience of use and, thus, the popularity of mobile devices for commercial transactions.

It is not easy to predict innovations in commercial transactions; the most revolutionary and successful innovations are the most difficult to predict, because they deviate from current practice. In particular, mobile multimedia devices can be expected to have major, but unforeseeable, effects on social interactions between people, as individuals and in groups. Novel forms of social interaction will inevitably engender new forms of commercial transactions.

REFERENCES

Bharat, R., & Minakakis, L. (2003). Evolution of mobile location-based services. *Communications of the ACM, 46*(12), 61-65.

Buranatrived, J., & Vickers, P. (2002). An investigation of the impact of mobile phone and PDA interfaces on the usability of mobile-commerce applications. *Proceedings of the IEEE 5th International Workshop on Networked Appliances*, Liverpool, UK.

Chung-wei, L., Wen-Chen, H., & Jyh-haw, Y. (2003). A system model for mobile commerce. *Proceedings of the IEEE 23rd International Conference on Distributed Computing Systems Workshops*, Providence, Rhode Island.

Eunseok, L., & Jionghua, J. (2003). A next generation intelligent mobile commerce system. *Proceedings of the ACIS 1st International Conference on Software Engineering Research and Applications*, San Francisco, California.

Hanebeck, H.C.L., & Raisinghani, M.S. (2002). Mobile commerce: Transforming vision into reality. *Journal of Internet Commerce, 1*(3), 49-64.

Jarvenpaa, S.L., Lang, K.R., Takeda, Y., & Tuunainen, V.K. (2003). Mobile commerce at crossroads. *Communications of the ACM, 46*(12), 41-44.

Keng, S., & Zixing, S. (2002). Mobile commerce applications, in supply chain management. *Journal of Internet Commerce, 1*(3), 3-14.

Kinoshita, M. (2002). DoCoMo's vision on mobile commerce. *Proceedings of the 2002 Symposium on Applications and the Internet*, Nara, Japan.

Kwok-Yan, L., Siu-Leung, C., Ming, G., & Jia-Guang, S. (2003). Lightweight security for mobile commerce transactions. *Computer Communications, 26*(18), 2052-2060.

Lane, M.S., Zou, Y., & Matsuda, T. (2002). NTT DoCoMo: A successful mobile commerce portal. *Proceedings of the 7th International Conference on Manufacturing and Management*, Bangkok, Thailand.

Lee, Y.E., & Benbasat, I. (2003). Interface design for mobile commerce. *Communications of the ACM, 46*(12), 48-52.

Ogawara S., Chen, J.C.H., & Chong P.P. (2002). Mobile commerce: The future vehicle of e-payment in Japan? *Journal of Internet Commerce, 1*(3), 29-41.

Ortiz, G.F., Branco, A.S.C., Sancho, P.R., & Castillo, J.L. (2002). ESTIA—Efficient electronic services for tourists in action. *Proceedings of the 3rd International Workshop for Technologies in E-Services*, Hong Kong, China.

Senn, J.A. (2000). The emergence of m-commerce. *IEEE Computer, 33*(12), 148-150.

Tarasewich, P. (2003). Designing mobile commerce applications. *Communications of the ACM, 46*(12), 57-60.

Urbaczewski, A., Valacich, J.S., & Jessup, L.M. (2003). Mobile commerce opportunities and challenges. *Communications of the ACM, 46*(12), 30-32.

Varshney, U., Vetter, R.J., & Kalakota, R. Mobile commerce: A new frontier. *IEEE Computer, 33*(10), 32-38.

WAP Forum. (2004). *http://www.wapforum.com*

Wen, H.J., & Gyires, T. (2002). The impact of wireless application protocol (WAP) on m-commerce security. *Journal of Internet Commerce, 1*(3), 15-27.

KEY TERMS

Cellular Communication: Wireless communication between mobile devices (e.g., cell phones, PDAs, and PCs) and fixed base stations. The base stations serve relatively small areas of a few square miles (called cells) and are interconnected by fixed telecommunication infrastructure that provides connection with other telecommunication systems. As a mobile device passes from one cell to another, one base station hands off the communication with the device to another without disrupting communication.

Mobile Commerce: Commercial transactions in which at least one party of the transaction uses a mobile wireless device, typically a cell phone, a PDA, or a PC equipped with a cellular modem. A PC can conduct a commercial Internet transaction using a WiFi connection to a base station, but because WiFi connections currently provide limited mobility, for this article, WiFi transactions are regarded as standard Internet transactions rather than mobile commerce.

Mobile Devices: Computing and communication devices, such as cell phones, PDAs, and PCs equipped with cellular modems. Mobile devices are lightweight and wireless so users can carry them and move about freely.

Mobile Multimedia: Use of audio and/or video in addition to text and image pages. The low bandwidth and high cost of mobile cellular connections discourage the use of video. Spoken natural language input and output is a promising but difficult approach for improving the ease of use of mobile devices for commercial transactions.

Wireless Application Protocol (WAP): The Wireless Application Protocol is an application-level communication protocol that is used to access services and information by hand-held devices with low-resolution displays and low bandwidth connections, such as mobile cell phones.

Wireless Markup Language (WML): A Web page description language derived from XML and HTML, but specifically designed to support the display of pages on low-resolution devices over low-bandwidth connections.

Wireless Transport Security Layer (WTSL): A high-security, low-overhead layer that operates above WDP and below WSP to provide authentication, authorization, confidentiality, integrity, and non-repudiation.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 638-644, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.6
Business Model Typology for Mobile Commerce

Volker Derballa
Universität Augsburg, Germany

Key Pousttchi
Universität Augsburg, Germany

Klaus Turowski
Universität Augsburg, Germany

ABSTRACT

Mobile technology enables enterprises to invent new business models by applying new forms of organization or offering new products and services. In order to assess these new business models, there is a need for a methodology that allows classifying mobile commerce business models according to their typical characteristics. For that purpose a business model typology is introduced. Doing so, building blocks in the form of generic business model types are identified, which can be combined to create concrete business models. The business model typology presented is conceptualized as generic as possible to be generally applicable, even to business models that are not known today.

INTRODUCTION

Having seen failures like WAP, the hype that was predominant for the area of mobile commerce (MC) up until the year 2001 has gone. About one year ago however, this negative trend has begun to change again. Based on more realistic expectations, the mobile access and use of data, applications and services is considered important by an increasing number of users. This trend becomes obvious in the light of the remarkable success of mobile communication devices. Substantial growth rates are expected in the next years, not only in the area of B2C but also for B2E and B2B. Along with that development go new challenges for the operators of mobile services resulting in re-assessed validations and alterations of existing business models and the creation of new business

models. In order to estimate the economic success of particular business models, a thorough analysis of those models is necessary. There is a need for an evaluation methodology in order to assess existing and future business models based on modern information and communication technologies. Technological capabilities have to be identified as well as benefits that users and producers of electronic offers can achieve when using them.

The work presented here is part of comprehensive research on mobile commerce (Turowski & Pousttchi, 2003). Closely related is a methodology for the qualitative assessment of electronic and mobile business models (Bazijanec, Pousttchi, & Turowski, 2004). In that work, the focus is on the added value for which the customers is ready to pay. The theory of informational added values is extended by the definition of technology-specific properties that are advantageous when using them to build up business models or other solutions based on information and communication techniques. As mobile communication techniques extend Internet technologies and add some more characteristics that can be considered as additional benefits, a own class of technology-specific added values is defined and named mobile added values (MAV), which are the cause of informational added values. These added values based on mobility of mobile devices are then used to assess mobile business models.

In order to be able to qualitatively assess mobile business models, those business models need to be unambiguously identified. For that purpose, we introduce in this chapter a business model typology. Further, the business model typology presented here is conceptualized as generic as possible, in order to be robust and be generally applicable — even to business models that are not known today. In the following we are building the foundation for the discussion of the business model typology by defining our view of MC. After that, alternative business model typologies are presented and distinguished from our approach, which is introduced in the subsequent section. The proposed approach is then used on an existing MC business model. The chapter ends with a conclusion and implications for further research.

BACKGROUND AND RELATED WORK

Mobile Commerce: A Definition

Before addressing the business model typology for MC, our understanding of MC needs to be defined. If one does agree with the Global Mobile Commerce Forum, mobile commerce can be defined as "the delivery of electronic commerce capabilities directly into the consumer's device, anywhere, anytime via wireless networks." Although this is no precise definition yet, the underlying idea becomes clear. Mobile commerce is considered a specific characteristic of electronic commerce and as such comprises specific attributes, as for example the utilization of wireless communication and mobile devices. Thus, mobile commerce can be defined as every form of business transaction in which the participants use mobile electronic communication techniques in connection with mobile devices for initiation, agreement or the provision of services. The concept mobile electronic communication techniques is used for different forms of wireless communication. That includes foremost cellular radio, but also technologies like wireless LAN, Bluetooth or infrared communication. We use the term mobile devices for information and communication devices that have been developed for mobile use. Thus, the category of mobile devices encompasses a wide spectrum of appliances. Although the laptop is often included in the definition of mobile devices, we have reservations to include it here without precincts due to its special characteristics: It can be moved easily, but it is usually not used during that process. For that reason we argue that the laptop can only be seen to some extend as amobile device.

Related Work

Every business model has to prove that it is able to generate a benefit for the customers. This is especially true for businesses that offer their products or services in the area of EC and MC. Since the beginning of Internet business in the mid 1990s, models have been developed that tried to explain advantages that arose from electronic offers. An extensive overview of approaches can be found in (Pateli & Giaglis, 2002). At first, models were rather a collection of the few business models that had already proven to be able to generate a revenue stream (Fedewa, 1996; Schlachter, 1995; Timmers, 1998). Later approaches extended these collections to a comprehensive taxonomy of business models observable on the web (Rappa, 2004; Tapscott, Lowi, & Ticoll, 2000). Only Timmers (1998) provided a first classification of eleven business models along two dimensions: innovation and functional integration. Due to many different aspects that have to be considered when comparing business models, some authors introduced taxonomies with different views on Internet business. This provides an overall picture of a firm doing Internet business (Osterwalder, 2002), where the views are discussed separately (Afuah & Tucci, 2001; Bartelt & Lamersdorf, 2000; Hamel, 2000; Rayport & Jaworski, 2001; Wirtz & Kleineicken, 2000). Views are for example commerce strategy, organizational structure or business process. The two most important views that can be found in every approach are value proposition and revenue. A comparison of views proposed in different approaches can be found in (Schwickert, 2004). While the view revenue describes the rather short-term monetary aspect of a business model the value proposition characterizes the type of business that is the basis of any revenue stream. To describe this value proposition authors decomposed business models into their atomic elements (Mahadevan, 2000). These elements represent offered services or products. Models that follow this approach are for example (Afuah & Tucci, 2001) and (Wirtz & Kleineicken, 2000). Another approach that already focuses on generated value can be found in (Mahadevan, 2000). There, four so-called value streams are identified: virtual communities, reduction of transaction costs, gainful exploitation of information asymmetry, and a value added marketing process.

In this work however, we are pursuing another approach: The evaluation of real business models showed that some few business model types recur. These basic business model types have been used for building up more complex business models. They can be classified according to the type of product or service offered. A categorization based on this criterion is highly extensible and thus very generic (Turowski & Pousttchi, 2003). Unlike the classifications of electronic offers introduced previously, this approach can also be applied to mobile business models that use for example location-based services to provide a user context. In the following sections, we are describing this business model typology in detail.

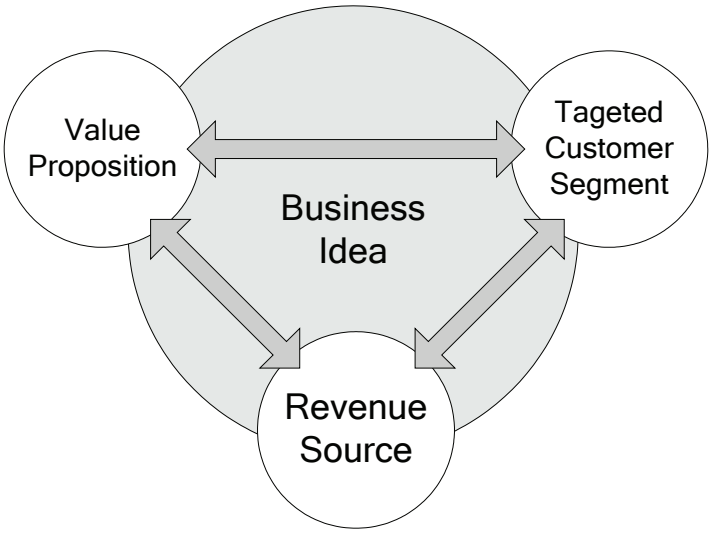
BUSINESS MODEL TYPOLOGY

Business Idea

Starting point for every value creation process is a product or business idea. An instance of a business idea is the offer to participate in auctions or conduct auctions — using any mobile device without tempo-spatial restrictions. Precondition for the economic, organisational, and technical implementation and assessment of that idea is its transparent specification. That abstracting specification of a business idea's functionality is called business model. It foremost includes an answer to the question: Why has this idea the potential to be successful? The following aspects have to be considered for that purpose:

- Value proposition (which value can be created)

Figure 1. Business idea and business model



- Targeted customer segment (which customers can and should be addressed)
- Revenue source (who, how much and in which manner will pay for the offer)

Figure 1 shows the interrelationship between those concepts. It needs to be assessed how the business idea can be implemented regarding organisational, technical, legal, and investment-related issues. Further, it has to be verified whether the combination of value proposition, targeted customer segment and revenue source that is considered optimal for the business model fits the particular company's competitive strategy. Let's assume an enterprise is pursuing a cost leader strategy using offers based on SMS, it is unclear whether the enterprise can be successful with premium-SMS.

It needs to be pointed out that different business models can exist for every single business idea. Coming back to the example of offering auctions without tempo-spatial restrictions, revenues can be generated in different ways with one business model recurring to revenues generated by advertisements and the other recurring to revenues generated by fees.

Revenue Models

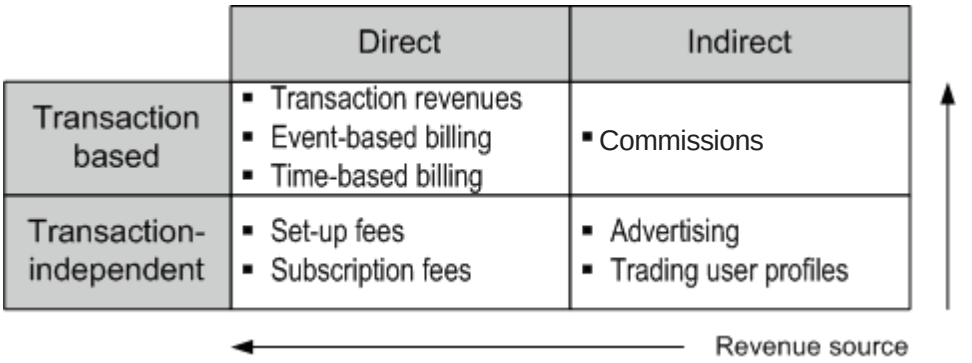
The instance introduced previously used the mode of revenue generation in order to distinguish business models. In this case, the revenue model is defined as the part of the business model describing the revenue sources, their volume and their distribution. In general, revenues can be generated by using the following revenue sources:

- Direct revenues from the user of a MC-offer
- Indirect revenues, in respect to the user of the MC-offer (i.e., revenues generated by 3rd parties); and
- Indirect revenues, in respect to the MC-offer (i.e., in the context of a non-EC offer).

Further, revenues can be distinguished according to their underlying mode in transaction-based and transaction-independent. The resulting revenue matrix is depicted in Figure 2.

Direct transaction-based revenues can include event-based billing (e.g., for file download) or time-based billing (e.g., for the participation in a blind-date game). Direct transaction-independent

Figure 2. Revenue sources in MC (based on Wirtz & Kleineicken, 2000)



revenues are generated as set-up fees, (e.g., to cover administrative costs for the first-time registration to a friend finder service) or subscription fees (e.g., for streaming audio offers).

The different revenue modes as well as the individual revenue sources are not necessarily mutually excluding. Rather, the provider is able to decide which aspects of the revenue matrix he wants to refer to. In the context of MC-offers, revenues are generated that are considered (relating to the user) indirect revenues. That refers to payments of third parties, which in turn can be transaction-based or transaction-independent. Transaction-based revenues (e.g., as commissions) accrue if, for example, restaurants or hotels pay a certain amount to the operator of mobile tourist guide for guiding a customer to their locality. Transaction-independent revenues are generated by advertisements or trading user profiles. Especially the latter revenue source should not be neglected, as the operator of a MC-offer possesses considerable possibilities for the generation of user profiles due to the inherent characteristics of context sensitivity and identifying functions (compared to the ordinary EC-vendor). Revenues that are not generated by the actual MC-offer are a further specificity of indirect revenues. This includes MC-offers pertaining to customer retention, effecting on other business activities (e.g., free SMS-information on a soccer team leading to an improvement in merchandising sales).

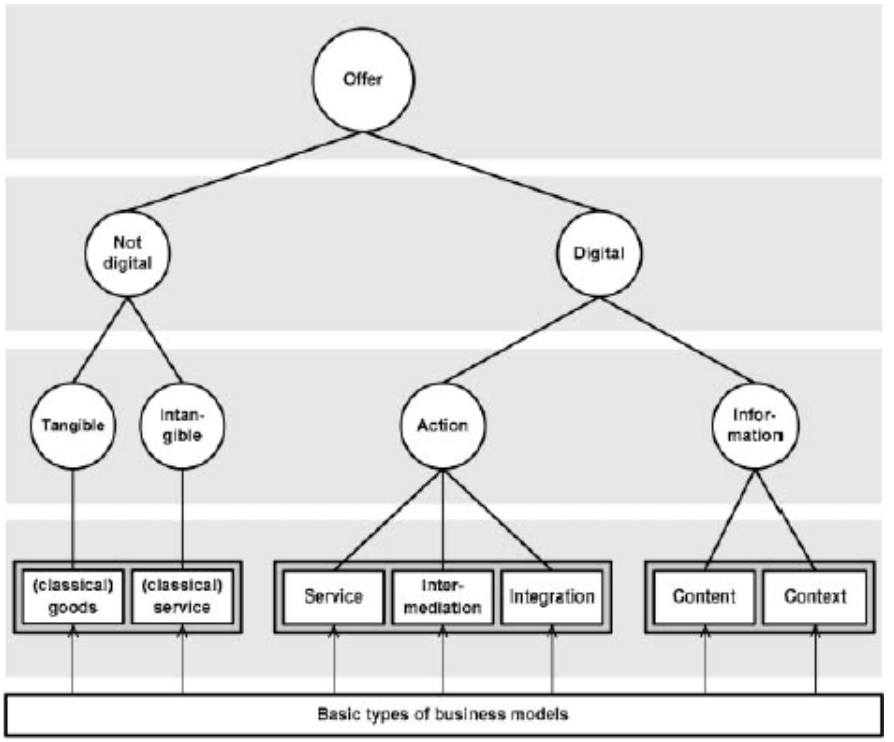
MC-Business Models

In the first step, the specificity of the value offered is evaluated. Is the service exclusively based on the exchange of digitally encoded data or is a significant not digital part existent (i.e., a good needs to be manufactured or a service is accomplished that demands some kind of manipulation conducted on a physically existing object)? *Not digital* services can be subdivided into *tangible* and *intangible* services. Whereas *tangible* services need to have a significant physical component, this classification assumes the following: The category of *intangible* services only includes services that demand manipulation conducted on a physically existing object.

Services that can be created through the exchange of digitally encoded data are subdivided into *action* and *information*. The category *information* focuses on the provision of data (e.g., multi-media contents in the area of entertainment or the supply of information). Opposed to that, the category *action* includes processing, manipulating, transforming, extracting, or arranging of data.

On the lowermost level, building blocks for business models are created through the further subdivision according to the value offered. For that purpose, a distinction is made between the concrete business models that can include one or more business model types and those business

Figure 3. Categorization of basic business model types



model types as such. These act as building blocks that can constitute concrete business models.

The business model type *classical goods* is included in all concrete business models aiming at the vending of tangible goods (e.g., CDs or flowers, i.e., goods that are manufactured as industrial products or created as agricultural produce). Those goods can include some digital components (e.g., cars, washing machines). However, decision criterion in that case is the fact that a significant part of the good is of physical nature and requires the physical transfer from one owner to the other.

Concrete business models include the business model type *classical service* if some manipulation activities have to be conducted on a physical object. That comprises e.g. vacation trips and maintenance activities.

The basic business model type *service* comprises concrete business models, if they comprise an original service that is considered by the customer as such and requires some action based on digitally encoded data as described previously, without having intermediation characteristics (c.f., basic business model type *intermediation*). Such services, e.g., route planning or mobile brokerage are discrete services and can be combined to new services through bundling. A typical offer that belongs to the business model type service is mobile banking. Further, it might be required (e.g., in order to enable mobile payment or ensure particular security goals (data confidentiality) to add further services, which require some kind of action, as described previously. As the emphasis is on the original service, these services can be considered as supporting factors. Depending on the circumstances, they might be seen as an original service. Due to that, those supporting services will not be attributed to a basic business model type. Rather, those services are assigned

to the business model type service.

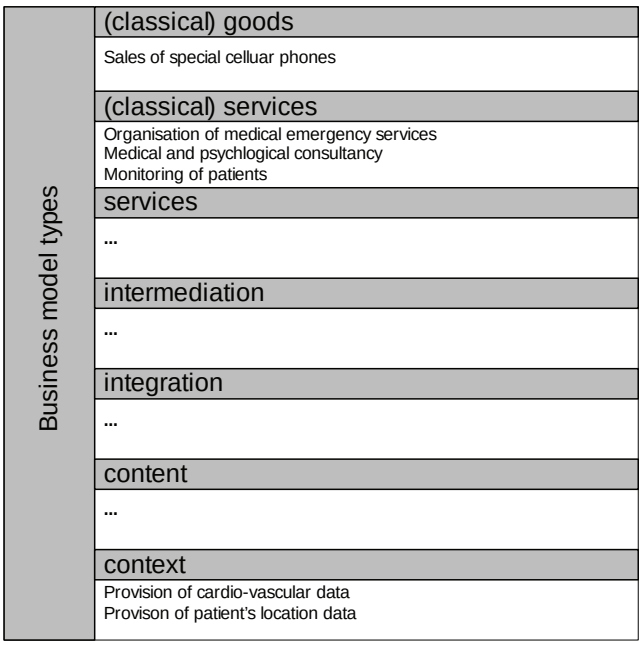
A concrete business model includes the business model type *intermediation* if it aims at the execution of classifying, systemising, searching, selecting, or interceding actions. The following offers are included:

- Typical search engines/offers (e.g., www.a1.net);
- Offers for detecting and interacting with other consumers demanding similar products;
- Offers for detecting and interacting with persons having similar interests;
- Offers for the intermediation of consumers and suppliers;
- Any kind of intermediation or brokering action, especially the execution of online auctions; and
- In general the operations of platforms (portals), which advance, simplify or enable the interaction of the aforementioned economic entities.

Taking all together, the focus is on matching of appropriate parings (i.e., the initiation of a transaction). Nevertheless, some offers provide more functionality by for example supporting the agreement process as well (e.g., the hotel finder and reservation service (wap.hotelkatalog.de)): This service lets the user search for hotels, make room reservations, and cancel reservations. All the relevant data is shown and hotel rooms can be booked, cancelled, or reserved. The user is contacted using e-mail, telephone, fax, or mail. Revenues are generated indirectly and transaction-independent, as the user agrees to obtain advertisements from third parties.

The basic business model type *integration* comprises concrete business models aiming at the combination of (original) services in order to create a bundle of services. The individual services might be a product of concrete business models that in turn can be combined to create new offers. Further, the fact that services have been combined is not necessarily transparent for the consumer. This can even lead to user individual

Figure 4. Classification of Vitaphone's business model



offers where the user does not even know about the combination of different offers. For example, an offer could be an insurance bundle specifically adjusted to a customer's needs. The individual products may come from different insurance companies. On the other hand, it is possible to present this combination to the consumer as the result of a customization process (custom-made service bundle).

The basic business model type *content* can be identified in every concrete business model that generates and offers digitally encoded multi-media content in the areas of entertainment, education, arts, culture sport etc. Additionally, this type comprises games. Wetter-Online (pda.wetteronline.de) can be considered a typical example for that business model type. The user can access free weather information using a PDA. The information offered includes forecasts, actual weather data, and holiday weather. The PDA-version of this service generates no revenues, as it is used as promotion for a similar EC-offer, which in turn is ad sponsored.

A concrete business model comprises the basic business model type *context* if information describing the context (i.e., situation, position, environments, needs, requirements etc.) of a user is utilised or provided. For example, every business model building on location-based services comprises or utilises typical services of the basic business model type context. This is also termed context-sensitivity. A multiplicity of further applications is realised in connection with the utilisation of sensor technology integrated in or directly connected to the mobile device. An instance is the offer of Vitaphone (www.vitaphone.de). It makes it possible to permanently monitor the cardiovascular system of endangered patients. In case of an emergency, prompt assistance can be provided. Using a specially developed mobile phone, biological signals, biochemical parameters, and the users' position are transmitted to the Vitaphone service centre. Additionally to the aforementioned sensors, the mobile phone has GPS functionality and a special emergency button to establish quick contact with the service centre.

Figure 4 depicts the classification of that business model using the systematics introduced previously. It shows that vita phone's business model uses mainly the building blocks from the area of *classical service*. Those services are supplemented with additional building blocks from the area of *context*. This leads to the weakening of the essential requirement — physical proximity of patient and medical practioner — at least what the medical monitoring is concerned. This creates several added values for the patient, which will lead to the willingness to accept that offer.

Analysing the offer of Vitaphone in more detail leads to the conclusion that the current offer is only a first step. The offer results indeed in increased freedom of movement, but requires active participation of the patient. He has to operate the monitoring process and actively transmit the generated data to the service centre. To round of the analysis of Vitaphone's business model, the revenue model is presented in Figure 5.

Non MC-relevant revues are generated by selling special cellular phones. Further, direct MC revenues are generated by subscription fees (with or without the utilisation of the service centre) and transmission fees (for data generated and telephone calls using the emergency button).

CONCLUSION

This chapter presents an approach to classify mobile business models by introducing a generic mobile business model typology. The aim was to create a typology that is as generic as possible, in order to be robust and applicable for business models that do not exist today. The specific characteristics of MC make it appropriate to classify the business models according to the mode of the service offered. Doing so, building blocks in the form of business model types can be identified. Those business model types then can be

Figure 5. Vitaphone's revenue model

MC-Business			Non-MC-Business
	Direct	Indirect	
Transaction based	▪ Communication with the service centre	...	Sales of special celluar phones
Transaction-independent	▪ Subscription fee	...	

combined to create concrete business model. The resulting tree of building blocks for MC business models differentiates digital and not digital services. Not digital services can be subdivided into the business model types classical goods for tangible services and classical service for intangible services. Digital services are divided into the category action with the business model types service, intermediation, integration and the category information with the business model types content and context.

Although the typology is generic and is based on the analysis of a very large number of actual business models, further research is necessary to validate this claim for new business models from time to time.

REFERENCES

Afuah, A., & Tucci, C. (2001). *Internet business models and strategies.* Boston: McGraw Hill.

Bartelt, A., & Lamersdorf, W. (2000). *Geschäftsmodelle des Electronic Commerce: Modell-bildung und Klassifikation.* Paper presented at the Verbundtagung Wirtschaftsinformatik.

Bazijanec, B., Pousttchi, K., & Turowski, K. (2004). *An approach for assessment of electronic offers.* Paper presented at the FORTE 2004, Toledo.

Fedewa, C. S. (1996). *Business models for Internetpreneurs.* Retrieved from http://www.gen.com/iess/articles/art4.html

Hamel, G. (2000). *Leading the revolution.* Boston: Harvard Business School Press.

Mahadevan, B. (2000). Business models for Internet based e-commerce: An anatomy. *California Management Review, 42*(4), 55-69.

Osterwalder, A. (2002). *An e-business model ontology for the creation of new management software tools and IS requirement engineering.* CAiSE 2002 Doctoral Consortium, Toronto.

Pateli, A., & Giaglis, G. M. (2002). *A domain area report on business models.* Athens, Greece: Athens University of Economics and Business.

Rappa, M. (2004). *Managing the digital enterprise — Business models on the Web.* Retrieved June 14, 2004, from http://digitalenterprise.org/models/models.html

Rayport, J. F., & Jaworski, B. J. (2001). *E-Commerce.* New York: McGraw Hill/Irwin.

Schlachter, E. (1995). *Generating revenues from Web sites.* Retrieved from http://boardwatch.internet.com/mag/95/jul/bwm39

Schwickert, A. C. (2004). *Geschäftsmodelle im electronic business—Bestandsaufnahme*

und relativierung. Gießen: Professur BWL-Wirtschaftsinformatik, Justus-Liebig-Universität.

Tapscott, D., Lowi, A., & Ticoll, D. (2000). *Digital capital—Harnessing the power of business Webs.* Boston.

Timmers, P. (1998). Business models for electronic markets. *Electronic Markets, 8,* 3-8.

Turowski, K., & Pousttchi, K. (2003). *Mobile Commerce—Grundlagen und Techniken*. Heidelberg: Springer Verlag.

Wirtz, B., & Kleineicken, A. (2000). Geschäftsmodelltypen im Internet. *WiSt, 29*(11), 628-636.

KEY TERMS

Business Model: Business model is defined as the abstracting description of the functionality of a business idea, focusing on the value proposition, customer segmentation, and revenue source.

Business Model Types: Building blocks for the creation of concrete business models.

Electronic Commerce: Every form of business transaction in which the participants use electronic communication techniques for initiation, agreement or the provision of services.

Mobile Commerce: Every form of business transaction in which the participants use mobile electronic communication techniques in connection with mobile devices for initiation, agreement or the provision of services.

Revenue Model: The part of the business model describing the revenue sources, their volume and their distribution.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 114-121,

Chapter 6.7
Making Money with Open-Source Business Initiatives

Paul Benjamin Lowry
Brigham Young University, USA

Akshay Grover
Brigham Young University, USA

Chris Madsen
Brigham Young University, USA

Jeff Larkin
Brigham Young University, USA

William Robins
Brigham Young University, USA

INTRODUCTION

Open-source software (OSS) is software that can be used freely in the public domain but is often copyrighted by the original authors under an open-source license such as the GNU General Public License (GPL). Given its free nature, one might believe that OSS is inherently inferior to proprietary software, yet this often is not the case. Many OSS applications are superior or on par with their proprietary competitors (e.g., MySQL, Apache Server, Linux, and Star Office). OSS is a potentially disruptive technology (Christensen, 1997) because it is often cheaper, more reliable, simpler, and more convenient than proprietary software.

Because OSS can be of high quality and capable of performing mission-critical tasks, it is becoming common in industry; the majority of Web sites, for example, use Apache as the Web

server. The deployment of OSS is proving to be a productive way to counter the licensing fees charged by proprietary software companies. An organized approach to distributing cost-effective OSS products is intensifying as companies such as RedHat and IBM co-brand OSS products to establish market presence.

From a business perspective, the entire OSS movement has been strategically anti-intuitive because it is based on software developers freely sharing source code—an act that flies in the face of traditional proprietary models. This movement raises two questions this article aims to address: (1) why would individuals write software and share it freely? and (2) how can software firms make money from OSS? Before fully addressing these questions, this article examines the historical development of OSS.

OSS HISTORY

A strategic irony of the software industry is that its foundation rests primarily on OSS principles. Software development in the 1960s and 1970s was steered primarily by government and academia. Software developers working in the field at the time considered it a normal part of their research culture to exchange, modify, and build on one another's software (Von Krogh, 2003). Richard Stallman, a professor and programmer at MIT, was a strong advocate and contributor to this culture of open, collaborative software development. Despite Professor Stallman's influence, MIT eventually stopped exchanging sourcecode with other universities to increase its research funding through proprietary software licensing. Offended by MIT's decision to limit code sharing, Professor Stallman founded the Free Software Foundation in 1985 and developed the General Public License (GPL) to preserve free code sharing (Bretthauer, 2002).

In the formative years of the software industry, Stallman's free software movement grew slowly; in the early 1990s, however, the concept of code sharing grew more rapidly for a couple of reasons. First, "free software" was renamed "OSS," a name that spread rapidly throughout the code-sharing community (Fitzgerald & Feller, 2001). Second, the OSS movement received a boost from the advent of the World Wide Web (WWW). The Web provided an opportunity for Internet users to quickly and conveniently share their code.

WHY DEVELOPERS WRITE OSS

The majority of OSS software developers fall into one of the following three categories: freelancers, software enthusiasts, or professionals. Freelancers

Table 1. Developer motivations

Enthusiast	Freelancer	Professional
• Learn	•Challenge of developing code	• Programming income
•Earn respect	•Receive future job opportunities	•Customize OSS

Intrinsic → Extrinsic

enjoy the challenges associated with developing OSS and providing services to the OSS community to further their own careers. When freelancers create modules of code, they often include their contact information inside the modules (Lerner & Tirole, 2002). This allows businesses to contact the developers to request their future services.

Software enthusiasts are people who contribute to OSS simply out of the joy and challenge of doing so, with little regard for professional advancement. Enthusiasts are often university students who want to participate in the development of free software and who receive personal gratification from participating in real-world OSS development projects and gaining the respect of the OSS community.

Even though OSS is "free" software, many companies hire professional developers to work on improving OSS code. RedHat, a Linux support company, hires developers to fix bugs in OSS code and to create new applications (Lerner & Tirole, 2002). Other companies hire OSS developers because their systems run OSS applications and they need developers to customize the code for specific business purposes. Table 1 summarizes the different motivations for joining OSS projects and shows them on a spectrum of intrinsic and extrinsic motivations.

SOFTWARE DEVELOPMENT ECONOMICS

Proprietary Software

The strategic motivation behind the creation of proprietary software is to set up high switching costs for consumers. For such companies their developers' resulting source code becomes the company's intellectual property and an unshared key company asset. Once customers purchase proprietary software, they must pay for updates continually to keep the software current, and often to receive full customer support (Delong & Froomkin, 2000). Most customers will pay these fees because of the lock in that occurs from the often costly prohibitive tradeoff of implementing a completely new system.

Microsoft is an example of a company that has succeeded in proprietary software, largely because they have a focused strategy of selling complementary products and services to their installed base of Windows users (Shapiro & Varian, 1998): Offering complementary goods that run on Windows (e.g., Office) increases profitability and successfully enhances the buyer relationship while encouraging customer entrenchment.

Proprietary software development is rigidly structured. Development begins with an end product in mind, and the new product often integrates with other products the company is currently selling. Project leaders create development plans, set deadlines, and coordinate teams to develop modules of the new software product. Successful proprietary software companies are also able to develop new technologies in exceptionally short time frames and to place their products in the market faster than their competitors. Products that meet the strict demands of end users succeed and increase customer satisfaction.

The downside of proprietary software development is that it comes at a tremendous internal cost (Lederer & Prasad, 1993); meanwhile, the industry is experiencing increasing pressures to decrease costs. Companies must invest heavily in research and development (R&D), human capital, information technology, marketing, brand development, and physical manufacturing of the products. They must continually innovate and develop updated versions of existing products, or create entirely new products. To compensate for these costs, proprietary software companies have high-priced products. Some software costs are so high that many businesses question whether the software is worth it.

OSS

The economics of OSS differ significantly in that OSS is developed in a loose marketplace structure. The development process begins when a developer presents an idea or identifies a need for an application with specific functionality (Johnson, 2002). OSS software development typically has a central person or body that selects a subset of developed code for an "official" release and makes it widely available for distribution. OSS is built by potentially large numbers of volunteers in combination with for-profit participants (Von Krogh, 2003). Often no system-level design or even detailed design exists. Developers work in arbitrary locations, rarely or never meet face to face, and often coordinate their activity through e-mail and bulletin boards. As participants make changes to the original application, the central person or body leading the development selects code changes, incorporates them into the application, and officially releases the next version of the application. Table 2 compares OSS to proprietary development.

OSS BUSINESS MODELS

A business model is a method whereby a firm builds and uses resources to provide a value-added proposition to potential customers (Afuah & Tucci, 2000). OSS business models are based on providing varied services that cater to cost-sensitive market segments and provide value to the end user by keeping the total cost of ownership as low as possible (Hecker, 1999). OSS-based companies must provide value-added services that are in demand, and they must provide these services at cost-sensitive levels. OSS is a strategic threat to proprietary software, because one of the most effective ways to compete in lock-in markets is to "change the game" by expanding the set of complementary products beyond those offered by rivals (Shapiro & Varian, 1998). OSS proponents are trying to "change the game" with new applications of the following business models (Castelluccio, 2000): support sellers, loss leaders, code developers, accessorizers, certifiers, and tracking service providers.

Table 2. OSS development vs. proprietary development

OSS	Proprietary Software
Similarities	
Building brand name and reputation increases software use	
Revenue is generated from supporting software, creating new applications for software, and certifying software users	
Differences	
Code developed outside of company for free	Developers are paid to program code
Source code is open for public use.	Source code is kept in company
People use program without paying any license fees.	Users pay license fees to use the software
Updates are free and users are allowed flexibility in using them	People are locked in using specific software and have to pay for updates
Code is developed for little internal cost	Code is costly to create internally

Support Sellers

Support sellers provide OSS to customers for free, except for a nominal packaging and shipping fee, and instead charge for training and consulting services. They also maintain the distribution channel and branding of a given OSS package. They provide value by helping corporations and individuals install, use, and maintain OSS applications. An example of a support seller is RedHat, which provides reliable Linux solutions.

To offer such services, support sellers must anticipate and provide services that will meet the needs of businesses using OSS. To offer reliable and useful consulting services, support sellers must invest heavily in understanding the currently available OSS packages and developing models to predict how these OSS applications will evolve in the future (Krishnamurthy, 2003).

This model has strengths in meeting the needs for outsourcing required IT services, which is the current market trend (Lung Hui & Yan Tam, 2002). OSS provides companies an opportunity to reduce licensing costs by allowing companies to outsource the required IT support to support sellers. Likewise, the marketplace structure of OSS development adds significant uncertainty to the future of OSS applications. Risk-adverse companies often do not want to invest in specialized human capital, and support sellers help mitigate these risks.

One drawback of this model is that consulting companies often fall prey to economic downturns, during which potential clients reduce outsourcing to consultants. This cycle is compounded for the software industry, since a poor economy results in cost cutting and an eventual reduction in IT spending.

Loss leaders

Loss-leader companies write and license proprietary software that can run on OSS platforms (Castelluccio, 2000). An example of a loss leader is Netscape, which gives away its basic Web-browser software but then provides proprietary software or hardware to ensure compatibility and allow expanded functionality. The loss-leader business model adds value by providing applications to companies that have partially integrated OSS with their systems (Hecker, 1999). Companies often need specific business applications that are unavailable in the OSS community, or they desire proprietary applications but wish to avoid high platform-licensing costs.

To leverage the integration of OSS with proprietary software, loss leaders need to assemble a team of highly skilled developers, create an IT infrastructure, and develop licensable applications. The major costs of this business model arise from payroll expenses for a development team, R&D costs, marketing, and, to a lesser extent, patenting and manufacturing.

This model's strength is that it provides a solution for the lack of business applications circulating in the OSS community. The loss leader model fills the gap between simpler available OSS applications, such as word processors, and more complex applications that are unavailable in the OSS community.

A weakness of this model is the risk of disintermediation. As time passes and OSS coding continues to grow and expand, more robust and complex applications will be developed. However, the developers of these applications will have to cope with the speed and efficiency of proprietary software development.

Code Developers

The code development model addresses some of the limitations of the loss-leader model. Code development companies generate service revenue through on-demand development of OSS. If a firm cannot find an OSS package that meets its needs for an inventory management system, for example, the firm could contract with a code development company to the basic application (Johnson, 2002).

The code development company could then distribute this application to the OSS community and act as the development project's leader. The code development company would track the changes made to the basic source code by the OSS community and integrate those changes into its product. The company would periodically send its customers product updates based on changes accepted from the OSS community.

The necessary assets and associated costs required by this model are similar to those in the proprietary software model, including a team of programmers, IT infrastructure, and marketing. However, the code developer needs to develop only a basic application. Once the basic software is developed, the OSS community provides further add-ons and new features (Johnson, 2002), which decrease the R&D costs for the company acting as project leader. Yet the code development team needs to have the necessary IT infrastructure to lead the OSS community in the application's evolution, incorporate new code, and resubmit new versions to its customers.

This model's strength is its longevity. The code development model overcomes the risk of disintermediation by basing its revenue generation on initiating OSS applications and maintaining leadership over their evolution; it does not focus on privatizing the development and licensing of applications.

This model's weakness is the risk of creating an application of limited interest to the OSS community. A possible solution to this problem would be an offer from the company leading the development process to reward freelance developers for exceptional additions to the application's original code.

Accessorizers

Accessorizers companies add value by selling products related to OSS. Accessorizers provide a variety of different value-added services, from installing Linux OS on their clients' hardware to writing manuals and tutorials (Hecker, 1999; Krishnamurthy, 2003). For example, O'Reilly & Associates, Inc. writes manuals for OSS and produces downloadable copies of Perl, a programming language.

One strength of this model is that it provides the new manuals and tutorials that the constantly changing nature of the OSS market requires. Another strength is its self-perpetuating nature: as more manuals and tutorials are produced, more people will write and use OSS applications, increasing the need for more manuals and tutorials.

This models' weakness is the difficulty of staying current with the many trends with the OSS community. This difficulty creates the risk of investing in the wrong products or producing too much inventory that is quickly outdated.

Certifiers

Certifiers establish methods to train and certify students or professionals in an application. Certificate companies like CompTIA generate revenue through training programs, course materials, examination fees, and certification fees. These programs provide value to the individuals enrolled in the certification programs and businesses looking for specific skills (Krishnamurthy, 2003). Certification helps the OSS industry by creating benchmarks, expectations, and standards employers can use to evaluate and hire employees based on specific skill sets.

Certification has long-term profit potential since most certification programs require recertification every few years due to continuing education requirements. Businesses value certification programs because they are a cost-effective way to train employees on new technologies. Certifiers, who achieve first-mover advantage, become trendsetters for the entire industry, increasing barriers to entry into the certification arena.

One downside of this model is the significant startup costs. Certifiers need to find qualified in-

dividuals to create manuals, teach seminars, and write tests. Certifiers must also survey businesses to discern which parts of specific applications are most important, and which areas need the greatest focus during training. Certifiers also need to gain substantial credibility through marketing and critical mass or their tests have little value. Increasing company name recognition and building a reputation in the certification arena can be an expensive and long process.

This model also faces the threat of disintermediation. Historically, certification programs have evolved into not-for-profit organizations, such as the AICPA in accounting, or the ISO 9000 certification in operations. The threat of obsolescence is another major weakness. In the 1970s, FORTRAN or COBOL certification may have been important (Castelluccio, 2000), but they have since become obsolete. Certifiers specializing in certain applications must be constantly aware of the OSS innovation frontier and adjust their certification options appropriately.

Tracking Service Providers

The tracking-services business model generates revenue through the sale of services dedicated to tracking and updating OSS applications. For example, many companies have embraced Linux to cut costs; however, many of these same companies have found it difficult to maintain and upgrade Linux because of their lack of knowledge and resources. Tracking-services companies, like Sourceforge.net and FreshMeat.net, sell services to track recent additions, define source code alternatives, and facilitate easy transition of code to their customers' systems.

A strength of this model is its ability to keep costs low by automating the majority of the work involved in tracking while still charging substantial subscription and download fees. However, these services must have Web-based interfaces with user-friendly download options, and they also must develop human and technological capabilities that find recent updates and distinguish between available alternatives.

A weakness of this model is low barriers to entry. This information-services model can be replicated with a simple Web interface and by spending time on OSS discussion boards and postings, creating the possibility of such services becoming commoditized. Table 3 summarizes some of the differences between the OSS business models.

CONCLUSION

The market battle between OSS and proprietary software has just begun. This battle could be termed a battle of complementary goods and pricing. For example, the strategies between Microsoft and RedHat are similar in that they both need a large, established user base that is locked in and has access to a large array of complementary goods and services. The key differences in their strategies are in their software development process, software distribution, intellectual property ownership, and pricing of core products and software. It will be increasingly important for OSS companies to track the competitive response of proprietary companies in combating the increasing presence of OSS.

Moreover, the OSS movement has begun to make inroads into the governments in China, Brazil, Australia, India, and Europe. As whole governments adopt OSS the balance of power can shift away from proprietary providers. This also provides the opportunity to develop a sustainable business model that caters only to the government sector. Similarly, formulating business models for corporations and educational institutions may be another fruitful opportunity.

The recent government regulations associated with the Sarbanes-Oxley Act and other financial-reporting legislation are important trends. These regulations require significant research in the area of internal control reporting on OSS applications.

Table 3. OSS models

Business Model	Assets	Costs	Revenue Model
Support Sellers	Human capital, supporting infrastructure, contracts	Payroll, IT, marketing and brand development	Training, consulting
Loss Leaders	Human capital, supporting infrastructure, software	Payroll, IT cost, marketing and brand development, R&D, software manufacturing	Licenses
Accessorizers	Human capital, supporting infrastructure	Payroll, printing material machines, training, software	Book Sales
Code developers	Human capital, software - technology tracking, database	Payroll, IT, marketing (Corporations), marketing (Freelancers)	Corporations that pay for service
Certifiers	Human capital, IT, Certification program	Certification program development, payroll	Tests, certificates
Tracking-service providers	Human capital, Software-technology tracking, Databases	Payroll, IT, marketing (Corporations)	Corporations that pay for service

It is likely the collaborative and less proprietary nature of OSS could help with this reporting. If this reporting can be done with more assurance than provided by proprietary applications, OSS providers can gain further advantage.

REFERENCES

Afuah, A. & Tucci, C. (2000). *Internet business models and strategies: Text and cases.* McGraw-Hill Higher Education.

Bretthauer, D. (2002). Open source software: A history. *Information Technology & Libraries, 21*(1), 3-10.

Castelluccio, M. (2000). Can the enterprise run on free software? *Strategic Finance, 81*(9), 50-55.

Christensen, C.M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail.* Harvard Business School Press.

Delong, J.B. & Froomkin, A.M. (2000). Beating Microsoft at its own game. *Harvard Business Review, 78*(1), 159-164.

Fitzgerald, B. & Feller, J. (2001). Guest editorial on open source software: Investigating the software engineering, psychosocial and economic issues. *Information Systems Journal, 11*(4), 273-276.

Hecker, F. (1999). Setting up shop: The business of open-source software. *IEEE Software, 16*(1), 45-51.

Johnson, J.P. (2002). Open source software: Private provision of a public good. *Journal of Economics & Management Strategy, 11*(4), 637-662.

Krishnamurthy, S. (2003). A managerial overview of open source software. *Business Horizons, 46*(5), 47-56.

Lederer, A.L. & Prasad, J. (1993). Information systems software cost estimating: A current assessment. *Journal of Information Technology, 8*(1), 22-33.

Lerner, J. & Tirole, J. (2002). Some simple economics of open source. *Journal of Industrial Economics, 50*(2), 197-234.

Lung Hui, K. & Yan Tam, K. (2002). Software functionality: A game theoretic analysis. *Journal of Management Information Systems (JMIS), 19*(1), 151-184.

MacCormack, A. (2001). Product-development practices that work: How Internet companies build software. *MIT Sloan Management Review, 42*(2), 75-84.

Shapiro, C. & Varian, H.R. (1998). *Information rules: A strategic guide to the network economy.* Harvard Business School Press.

Von Krogh, G. (2003). Open-source software development. *MIT Sloan Management Review, 44*(3), 14-18.

KEY TERMS

Copyright: A legal term describing rights given to creators for their literary and artistic works. See World Intellectual Property Organization at www.wipo.int/about-ip/en/copyright.html.

General Public License (GPL): License designed so that people can freely (or for a charge) distribute copies of free software, receive the source code, change the source code, and use portions of the source code to create new free programs.

GNU: GNU is a recursive acronym for "GNU's Not Unix." The GNU Project was launched in 1984 to develop a free Unix-like operating system. See www.gnu.org/.

Open-source Software (OSS): Software that can be freely used in the public domain, but is often copyrighted by the original authors under an open-source license such as the GNU GPL. See the Open Source Initiative at www.opensource.org/docs/definition_plain.php.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by J. Wang, pp. 555-561, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.8
Learning through Business Games

Luigi Proserpio
Bocconi University, Italy

Massimo Magni
Bocconi University, Italy

BUSINESS GAMES: A NEW LEARNING TOOL

Managerial business games, defined as interactive computer-based simulations for managerial education, can be considered as a relatively new tool for adults' learning. If compared with paper-based case histories, they could be less consolidated in terms of design methodologies, usage suggestions, and results measurement.

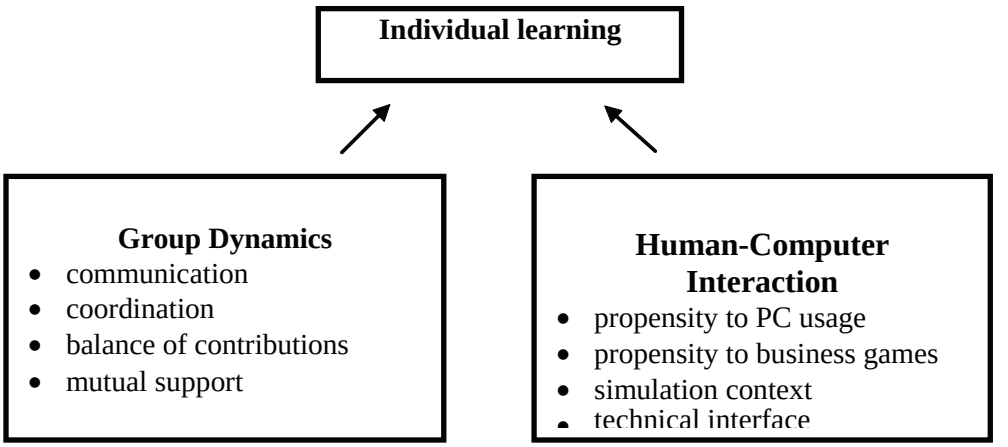
Due to the growing interest around Virtual Learning Environment (VLE), we are facing a positive trend in the adoption of business games for undergraduate and graduate education. This process can be traced back to two main factors. On the one hand, there is an increasing request for non-traditional education, side by side with an educational model based on class teaching (Alavi & Leidner, 2002). On the other hand, the rapid development of information technologies has made available specific technologies built around learning development needs (Webster & Hackley, 1997). Despite the increased interest generated by business games, many calls have still to be addressed on the design and utilization side. This contribution describes two fundamental aspects related with business games in graduate and undergraduate education: group dynamics (as current business games are almost in all instances played in groups) and human-computer interaction.

Figure 1 represents the variables that could influence individual learning in a business game context.

THE INFLUENCE OF GROUP DYNAMICS

It is widely accepted that a positive climate among subjects is fundamental to enhance the productivity of the learning process (Alavi, Wheeler,

Figure 1. Variables influencing individual learning in a business game setting



& Valacich, 1995). This is why group dynamics are believed to have a strong impact on learning within a team based context. A clear explanation of group dynamics impact on performance and learning is well developed in the teamwork quality construct (TWQ) (Hoegl & Gemuenden, 2001). Group relational dynamics are even more important when the group is asked to solve tasks requiring information exchange and social interaction (Gladstein, 1984), such as a business game. In fact, the impact of social relations is deeper when the task is complex and characterized by sequential or reciprocal interdependencies among members.

With reference to TWQ, it is possible to point out different group dynamics variables with a strong influence on individual learning in a business game environment: communication, coordination, balance of contributions, and mutual support. Instructors and business games designers should carefully consider the following variables, in order to maximize learning outcomes.

Hereafter, focusing on a business game setting, we will discuss each of these concepts and their relative impact on individual learning.

Communication

In order to develop effective group decision processes, information exchange among members should also be effective. In fact, communication is the way by which members exchange information (Pinto & Pinto, 1990), and smooth group functioning depends on communication easiness and efficacy among members (Shaw, 1981).

Moreover, individuals should be granted an environment where communication is open. A lack of openness should negatively influence the integration of knowledge and group members' experiences (Gladstein, 1984; Pinto & Pinto, 1990). These statements are confirmed by several empirical studies, showing direct and strong correlation between communication and group performance (Griffin & Hauser, 1992). According to Kolb's experiential learning theory, in a learning setting based on experiential methods (i.e., business game), it is important to provide the classroom with an in-depth debriefing in order to better understand the link between the simulation and the related theoretical assumption.

For these reason, groups with good communication dynamics tend to adopt a more participative behavior during the debriefing session, with higher quality observations. As a consequence, there is

a process improvement in the acquisition, generation, analysis and elaboration of information among members (Proserpio & Magni, 2004).

Balance of Contributions

It can be defined as the level of participation of each member in the group decision process. Each member, during the decision process, brings to the group a set of knowledge and experiences that allows the group to develop a cognitive advantage over individual decision process. Thus, it is necessary that each member brings his/her contributions to the group (Seers, Petty, & Cashman, 1995) in order to improve performance, learning and satisfaction of team members (Seers, 1989). A business game setting requires a good planning and implementation of strategies in order to better face the action-reaction process with the computer. For this reason, a balanced contribution among members favors the cross fertilization and the development of effective game strategies.

Coordination

A group could be seen as a complex entity integrating the various competencies required to solve a complex task. For this reason, a good balance of members' contribution is a necessary condition, although not sufficient. The expression of the group cognitive advantage is strictly tied to the harmony and synchronicity of members' contribution, that is, the degree according to which they coordinate their individual activities (Tannenbaum, Beard, & Salas, 1992).

As for communication, individuals belonging to groups with a better coordination level show better interventions in the debriefing phases. They also offer good hints to deepen the topics included in the simulation, playing as an intellectual stimulus for each other.

Mutual Support

It can be defined as the emergence of cooperative and mutually supporting behaviors, which lead to better team effectiveness (Tjosvold, 1984). In contrast, it is important to underline that competitive behaviors within a team determine distrust and frustration.

Mutual support among participants in a business game environment could be seen as an interference between the single user and the simulation: every discussion among users on simulation interpretation distracts participants from the ongoing simulation. This is why the emergence of cooperative behaviors does not univocally lead to more effective learning processes. These relations lower users' concentration and result in obstacles in the goal achievement path. Moreover, during a business game, users play in a time pressure setting, which brings to a drop in the effectiveness of the decision process. All these issues, according to group effectiveness theories, help to understand how mutual support in a computer simulation environment could show a controversial impact on individual learning (Proserpio & Magni, 2004).

THE INFLUENCE OF HUMAN-COMPUTER INTERACTION

Business games are often described as proficient learning tools. Despite the potentiality, as stressed by Eggleston and Janson (1997), there is the need for an in-depth analysis of the relationship between user and computer. On the design side, *naïve* business games (not designed by professionals) can hinder the global performance of a simulation and bring to negative effects on the learning side. For these reason, technological facets are considered as a fundamental issue for a proficient relationship between user and computer in order to improve learning process effectiveness (Alavi & Leidner, 2002; Leidner & Jarvenpaa, 1995).

Propensity to PC Usage

Attitude toward PC usage can be defined as the user's overall affective reaction when using a PC (Venkatesh et al., 2003). Propensity to PC usage can be traced back to the concepts of pleasure, joy, interest associated with technology usage (Compeau, Higgins, & Huff, 1999). It is consistent to think that users' attitude towards computer use could influence their use involvement, increasing or decreasing the impact of simulation on learning process.

From another standpoint, more related to HCI theories, computer attitude is tied to the simulation easiness of use. It is possible to argue that a simple simulation does not require strong computer attitude to enhance the leaning process. On the contrary, a complex simulation could worsen individual learning, because the cognitive effort of the participant can be deviated from the underlying theories to a cumbersome interface.

Propensity to Business Game Usage

This construct can be defined as the cognitive and affective elements that bring the user to assume positive/negative behaviors toward a business game. In fact, in these situations, users can develop feelings of joy, elation, pleasure, depression, or displeasure, which have an impact on the effectiveness of their learning process (Taylor & Todd 1995). Consistently with Kolb's theory (Kolb, 1984) on individual different learning styles, propensity to simulations could represent a very powerful element to explain individual learning.

Simulation Context

The simulation context can be traced back to the role assumed by individuals during the simulation. In particular, it is referred to the role of participants, teacher and their relationship. Theory and practice point out that business games have to be self-explaining. In other words, the intervention of other users or the explanations of a teacher to clarify simulation dynamics have to be limited. Otherwise, the user's effort to understand the technical and interface features of the simulation could have a negative influence on learning objectives (Whicker & Sigelman, 1991). Comparing this situation with a traditional paper-based case study, it is possible to argue that good instructions and quick suggestions during a paper-based case history analysis can help in generating users' commitment and learning. On the contrary, in a business game setting, a self explanatory simulation could bring users to consider the intervention of the teacher as an interruption rather than a suggestion. Thus, simulations have often an impact on the learning process through the reception step (Alavi & Leidner, 2002), meaning that teacher's or other members' intervention hinder participants to understand incoming information.

Technical Interface

Technical interface can be defined as the way in which information is presented on the screen (Lindgaard, 1994). In a business game, the interface concept is also related to the interactivity facet (Webster & Hackley, 1997). Several studies have pointed out the influence of technical interface on user performance and learning (Jarvenpaa, 1989; Todd & Benbasat, 1991). During the business game design, it is important to pay attention to the technical interface. It is essential that the interface captures user attention, thereby increasing the level of participation and involvement. According to the above mentioned studies, it is possible to argue that an attractive interface could represent one of the main variables that influence the learning process in a business game setting.

Face Validity

Face validity defines the coherence of simulation behaviors in relation with the user's expectancies

on perceived realism. It is also possible to point out that the perceived soundness of the simulation is a primary concept concerning the users' learning (Whicker & Sigelman, 1991). The simulation cannot react randomly to the user's stimulus, but it should recreate a certain logic path which starts from player action and finishes with the simulation reaction. It is consistent with HCI and learning theories, to argue that an effective business game has to be designed to allow users to recognize a strong coherence among simulation reactions, their actions, and their behavior expectancies.

ADDITIONAL ISSUES TO DESIGN A BUSINESS GAME

The main aspect that has to be considered when designing a business game is the ability of the simulation to create a safe test bed to learn management practices and concepts. It is fundamental that users are allowed to experiment behaviors related to theoretical concepts without any real risk. This issue, together with aspects of fun and the creation of a group collaboration context, could be useful to significantly improve the learning level.

Thus, a good simulation is based on homomorphic assumptions. Starting from the existence of a reality with *n* characteristics, homomorphism is the ability to choose *m* (with $n>m$) characteristics of this reality in order to reduce its complexity without losing too much relevant information. For example, in a F1 simulation game, racing cars can have a different behavior on a wet or dry circuit, but they cannot have a different behavior among wet, very wet, or almost wet.

In order to minimize the negative impact on learning processes, it is important that characteristics not included in the simulation should not impact too much on the simulation realism.

CONCLUSION

Several studies have shown the importance of involvement and participation in the fields of standard face-to-face education and in distance learning environments (Webster & Hackley, 1997). This research note extends the validity of previous statements to the business game field.

The discussion above allows us to point out a relevant impact on learning of two types of variables, while using a business game: group dynamics and human-computer interaction.

From previous researches, it is possible to argue that the "game" dimension captures a strong part of participants' cognitive energies (Proserpio & Magni, 2004). The simulation should be designed in a fashion as interactive as possible. Moreover, instructors should take into account that their role is to facilitate the simulation flow, leaving the game responsibility to transmit experiences on theories and their effects.

This is possible if the simulation is easy enough to understand and use. In this case, despite the fact that the simulation is computer based, there is not the emergence of a strong need for computer proficiency. This conclusion is consistent with other researches which showed the impact of the easiness of use on individual performance and learning (Delone & McLean, 1992).

The relationship between user and machine is mediated by the interface designed for the simulation, which represents a very powerful variable to explain and favor the learning process with these high involvement learning tools.

Computer simulations seem to have their major strength in the computer interaction, which ought to be the main focus in the design phase of the game. Interaction among groups' members is still important, but less relevant than the interface on individual learning.

REFERENCES

Alavi, M. & Leidner D. (2002). Virtual learning systems. In H. Bidgole (Ed.), *Encyclopedia of Information Systems* (pp. 561-572). Academic Press.

Alavi, M., Wheeler, B.C., & Valacich, J.S. (1995). Using IT to reengineer business education: An exploratory investigation of collaborative telelearning, *MIS Quarterly, 19*(3), 293-312.

Compeau, D.R., Higgins, C.A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly, 23*(2), 145-158.

Delone, W.H. & McLean, E.R. (1992). Information systems success: The quest for dependent variables. *Information Systems Research, 3*(1), 60-95.

Eggleston, R.G. & Janson, W.P. (1997). Field of view effects on a direct manipulation task in a virtual environment. *Proceedings of the Human Factors and Ergonomic Society 41st Annual Meeting*, (pp. 1244-1248).

Gladstein, D.L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29*, 499-517.

Griffin, A. & Hauser, J.R. (1992). Patterns of communication among marketing, engineering, and manufacturing: A comparison between two new product development teams. *Management Science, 38*(3), 360-373.

Hoegl, M. & Gemuenden, H.G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science, 12*(4), 435-449.

Institute of Electrical and Electronics Engineers (1990). *IEEE standard computer dictionary: A compilation of IEEE standard computer glossaries.* New York.

Jarvenpaa, S.L. (1989). The effect of task demands and graphical format on information processing strategies. *Management Science, 35*(3), 285-303.

Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall.

Leidner, D.E. & Jarvenpaa, S.L. (1995). The use of information technology to enhance management school education: A theoretical view. *MIS Quarterly, 19*(3), 265-292.

Lindgaard, G. (1994). *Usability testing and system evaluation: A guide for designing useful computer systems.* London; New York: Chapman & Hall.

Pinto, M.B. & Pinto, J.K. (1990). Project team communication and cross functional cooperation in new program development. *Journal of Product Innovation Management, 7*, 200-212.

Proserpio, L. & Magni, M. (2004). To play or not to play. Building a learning environment through computer simulations. *ECIS Proceedings*, Turku, Finland.

Seers, A., Petty, M., & Cashman, J.F., (1995). Team-member exchange under team and traditional management: A naturally occurring quasi experiment. *Group & Organization Management, 20*, 18-38.

Seers, A. (1989). Team-member exchange quality: A new construct for role-making research. *Organizational Behavior and Human Decision Process, 43*, 118-135.

Shaw, M.E. (1981). *Group dynamics: The psychology of small group behavior.* New York: McGraw-Hill.

Tannenbaum, S.I., Beard, R.L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. K. Kelley, (a cura di), *Issues, Theory, and Research in Industrial/Or-*

ganizational Psychology. Elsevier, Amsterdam, Holland, 117-153.

Taylor, S. & Todd, P.A. (1995). Assessing IT usage: The role of prior experience. *MIS Quarterly, 19*(2), 561-570.

Tjosvold, D. (1984). Cooperation theory and organizations. *Human Relations, 37*(9), 743-767.

Todd, P.A. & Benbasat, I. (1991). An experimental investigation of the impact of computer based decision aids on decision making strategies. *Information Systems Research, 2*(2), 87-115.

Venkatesh, V. et al. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478.

Webster, J. & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning, *Academy of Management Journal, 40*(6), 1282-1310.

Whicker, M.L. & Sigelman, L. (1991). *Computer simulation applications: An introduction.* Newbury Park, CA: Sage Publications.

Wight, A. (1970). Participative education and the inevitable revolution. *Journal of Creative Behavior, 4*(4), 234-282.

KEY TERMS

Business Games: Computer-based simulations designed to learn business-related concepts.

Experiential Learning: A learning model "which begins with the experience followed by reflection, discussion, analysis and evaluation of the experience" (Wight, 1970, p. 234-282).

HCI (Human-Computer Interaction): A scientific field concerning design, evaluation, and implementation of interactive computing systems for human usage.

Interface: An interface is a set of commands or menus through which a user communicates with a software program.

TWQ (Teamwork Quality): A comprehensive concept of the quality of interactions in teams. It represents how well team members collaborate or interact.

Usability: The ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component (Institute of Electrical and Electronics Engineers).

VLE (Virtual Learning Environments): Computer-based environments for learning purposes.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 532-537, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.9
Internet Privacy from the Individual and Business Perspectives

Tziporah Stern
Baruch College, CUNY, USA

INTRODUCTION: PRIVACY

People have always been concerned about protecting personal information and their right to privacy. It is an age-old concern that is not unique to the Internet. People are concerned with protecting their privacy in various environments, including healthcare, the workplace and e-commerce. However, advances in technology, the Internet, and community networking are bringing this issue to the forefront. With computerized personal data files:

a. Retrieval of specific records is more rapid;
b. Personal information can be integrated into a number of different data files; and
c. Copying, transporting, collecting, storing, and processing large amounts of information are easier.

In addition, new techniques (i.e., data mining) are being created to extract information from large databases and to analyze it from different perspectives to find patterns in data. This process creates new information from data that may have been meaningless, but in its new form may violate a person's right to privacy. Now, with the World Wide Web, the abundance of information available on the Internet, the many directories of information easily accessible, the ease of collecting and storing data, and the ease of conducting a search using a search engine, there are new causes for worry (Strauss & Rogerson, 2002). This article outlines the specific concerns of individuals, businesses, and those resulting from their interaction with each other; it also reviews some proposed solutions to the privacy issue.

CONTROL: PRIVACY FROM THE INDIVIDUAL'S PERSPECTIVE

The privacy issue is of concern to many types of people and individuals from different backgrounds. Gender, age, race, income, geographical location, occupation, and education level all affect people's views about privacy. In addition, culture (Milberg et al., 2000; Smith, 2001) and the amount of Web experience accumulated by an individual is likely to influence the nature of the information considered private (Hoffman et al., 1999; Miyazaki & Fernendez, 2001). Table 1 summarizes the kinds of information people would typically consider private.

When interacting with a Web site, individuals as consumers are now more wary about protecting their data. About three-quarters of consumers who are not generally concerned about privacy fear intrusions on the Internet (FTC, 2000). This is due to the digitalization of personal information, which makes it easier for unauthorized people to access and misuse it (see Table 2 for a list of concerns regarding the uses of data). For example, many databases use Social Security numbers as identifiers. With this information and the use of the Internet, personal records in every state's municipal database can be accessed (Berghel, 2000).

Table 1. Private information

Information
▪ Address
▪ Credit card numbers
▪ Date of birth
▪ Demographic information
▪ E-mail
▪ Healthcare information and medical records
▪ Name
▪ Phone number
▪ Real-time discussion
▪ Social Security number
▪ Usage tracking/click streams (cookies)

There also are many issues regarding policies and security controls. Individuals are concerned about breaches of security and a lack of internal controls (Hoffman, 2003). However, surprisingly, about one-third of Web sites do not post either a privacy policy or an information practice statement (Culnan, 1999), and only about 10% address all five areas of the Fair Information Practices (FIP), U.S. guidelines to protect computerized information (see FIP in Terms section) (Culnan, 1999; Federal Trade Commission, 2000). Additionally, there is a mismatch between policies and practices (Smith, 2001); this means that a company may publicize fair information policies but in practice does not follow its own guidelines.

Furthermore, as a result of the data mining technology, computer merging and computer matching have become a new privacy concern. One reason is because individuals may have authorized data for one purpose but not for another, and through data mining techniques, this information is extracted for further use and analysis. For example, a consumer's information may have been split up among many different databases. However, with sophisticated computer programs, this information is extracted and used to create a new database that contains a combination of all the aggregate information. Some of these data mining techniques may not be for the benefit of the consumer. It may allow the firms to engage in price and market discrimination by using consumers' private information against them (Danna & Gandy, 2002).

Some additional concerns are whether the Web site is run by a trusted organization, whether individuals can find out what information is stored about them, and whether their name will be removed from a mailing list, if requested. Consumers also want to know who has access to the data and if the data will be sold to or used by third parties. They want to know the kind of information collected and the purpose for which it is collected (Cranor et al., 1999; Hoffman, 2003). In addition, consumers want to feel in control

Table 2. Individual's concerns

Concerns
▪ Access
▪ Analyzing
▪ Collection
▪ Combining data
▪ Contents of the consumer's data storage device
▪ Creating marketing profiles of consumers
▪ Cross matching
▪ Distributing and sharing
▪ Errors in data
▪ Identity theft
▪ Reduced judgment in decision making
▪ Secondary use of data
▪ Selling data (government)
▪ Spam
▪ Storing
▪ Use
▪ Video surveillance on the Internet
▪ Web bugs

of their personal information (Hoffman, 2003; Olivero & Lunt, 2004). According to a Harris Poll (2003), 69% of consumers feel they have lost control of their personal information.

TRUST: PRIVACY FROM THE BUSINESS PERSPECTIVE

Privacy also is important to businesses. A business collects information about its customers for many reasons: to serve them more successfully, to build a long-term relationship with them, and to personalize services. To build a successful relationship, businesses must address their customers' privacy concerns (Resnick & Montania, 2003) so that their customers will trust them. They must also protect all information they have access to, since this is what consumers expect of them (Hoffman et al., 1999). Furthermore, they must be aware of the fact that some information is more sensitive (Cranor et al., 1999), such as Social Security numbers (Berghel, 2000). This trust is the key to building a valuable relationship with customers (Hoffman et al., 1999; Liu et al. 2004).

One of the many ways a business can gain consumer confidence is by establishing a privacy policy, which may help consumers trust it and lead them to return to the Web site to make more purchases (Liu et al., 2004). When a business is trusted, consumers' privacy concerns may be suppressed, and they may disclose more information (Xu et al., 2003). Privacy protection thus may be even more important than Web site design and content (Ranganathan & Ganapathy, 2002). Also, if an organization is open and honest with consumers, the latter can make a more informed decision as to whether or not to disclose information (Olivero & Lunt, 2004).

INDIVIDUAL VS. BUSINESS = PRIVACY VS. PERSONALIZATION

In matters of information, there are some areas of conflict between businesses and consumers. First, when a consumer and an organization complete a transaction, each has a different objective. The consumer does not want to disclose any personal information unnecessarily, and a business would like to collect as much information as possible about its customers so that it can personalize services and advertisements, target marketing efforts, and serve them more successfully. Consumers do appreciate these efforts yet are reluctant to share private information (Hoffman, 2003).

Cookies

Second, search engines also may potentially cause privacy problems by storing the search habits of their customers by using cookies. Their caches also may be a major privacy concern, since Web pages with private information posted by mistake, listserv, or Usenet postings may become available worldwide (Aljifri & Navarro, 2004). In general, cookies may be a privacy threat by saving personal information and recording user habits. The convenience of having preferences

saved does not outweigh the risks associated with allowing cookies access to your private data. There are now many software packages that aid consumers in choosing privacy preferences and blocking cookies (see solutions section).

Google

Finally, the most recent controversy involves Google's Gmail service and Phonebook. Gmail uses powerful search tools to scan its users' e-mails in order to provide them with personalized advertisements. On the one hand, this invades users' private e-mails. However, it is a voluntary service the user agrees to when signing up (Davies, 2004).

SOLUTIONS

There have been many attempts at trying to solve the privacy problem. There are three different types of solutions: governmental regulation, self-regulation, and technological approaches.

Governmental Regulation

Some form of government policy is essential, since in the absence of regulation and legislation to punish privacy-offenders, consumers may be reluctant to share information. However, written privacy policy requires enforcement (O'Brien & Yasnoff, 1999). In addition, given the current bureaucratic nature of legislation, technology advances far faster than the laws created to regulate it. Consequently, self-regulation may be a better solution.

Self-Regulation

There are numerous forms of self-regulation. The Fair Information Practices (U.S.) and the Organization for Economic Co-operation and Development (OECD) Guidelines (International) are both guidelines for protecting computerized records. These guidelines provide a list of policies a company should follow. Another type of self-regulated solution is a privacy seal program such as TRUSTe or Verisign. A business may earn these seals by following the guidelines that the seal company provides. A third kind of self-regulation is opt-in/opt-out policy. Consumers should be able to choose services by opting in or out (Hoffman et al., 1999) and to voluntarily embrace new privacy principles (Smith, 2001). A joint program of privacy policies and seals may provide protection comparable to government laws (Cranor et al., 1999) and may even address new issues faster than legislation.

Technology

Technological solutions also are a viable alternative. Technologies can protect individuals by using encryption, firewalls, spyware, and anonymous and pseudonymous communication. A well-known privacy technology is the Platform for Privacy Preferences (P3P), a World Wide Web Consortium (W3C) project that provides a framework for online interaction and assists users in making informed privacy decisions. In summary, although there seems to be some promise to each of these three alternatives, a combination of government regulation, privacy policies, and technology may be the best solution.

CONCLUSION

Advances in the collection and analysis of personal information have proven to be beneficial to society. At the same time, they have aggravated the innate concern for the protection of privacy. This article has reviewed current issues in the areas of information privacy and its preservation. It has included the differing points of view of those providing the information and those collecting and using it. Since the collection of information entails

both benefits and threats, various suggestions for minimizing the economic costs and maximizing the benefits are discussed.

REFERENCES

Aljifri, H., & Navarro, D.S. (2004). Search engines and privacy. *Computers and Security, 23*(5), 379-388.

Berghel, H. (2000). Identity theft, Social Security numbers, and the Web. *Communications of the ACM, 43*(2), 17-21.

Cranor, L.F., Reagle, J., & Ackerman, M.S. (1999). Beyond concern: Understanding net users' attitudes about online privacy. *AT&T Labs-Research Technical Report TR 99.4.3*. Retrieved April 5, 2004, from *http://www.research.att.com/resources/trs/TRs/99/99.4/99.4.3/report.htm*

Culnan, M.J. (1999). Georgetown Internet privacy policy survey: Report to the Federal Trade Commission. Retrieved April 3, 2004, from *http://www.msb.edu/faculty/culnanm/gippshome.html*

Danna, A., & Gandy Jr., O.H. (2002). All that glitters is not gold: Digging beneath the surface of data mining. *Journal of Business Ethics, 40*(4), 373-386.

Davies, S. (2004). Privacy international complaint: Google Inc.—Gmail email service. Retrieved June 22, 2004, from *http://www.privacyinternational.org/issues/internet/gmail-complaint.pdf*

Federal Trade Commission. (2000). Privacy online: Fair information practices in the electronic marketplace. Retrieved September 23, 2004, from *http://www.ftc.gov/reports/privacy2000/privacy2000.pdf*

Harris Poll. (2003). Most people are "privacy pragmatists" who, while concerned about privacy, will sometimes trade it off for other benefits. Retrieved September 23, 2004, from *http://www.harrisinteractive.com/harris_poll/index.asp?PID=365*

Hoffman, D.L. (2003). The consumer experience: A research agenda going forward. *FTC public workshop1: Technologies for protecting personal information: The consumer experience. Panel: Understanding how consumers interface with technologies designed to protect consumer information*. Retrieved June 6, 2004, from *http://elab.vanderbilt.edu/research/papers/pdf/manuscripts/FTC.privacy.pdf*

Hoffman, D.L., Novak, T.P., & Peralta, M. (1999). Information privacy in the marketplace: Implications for the commercial uses of anonymity on the Web. *The Information Society, 15*(2), 129-140.

Liu, C., Marchewka, J.T., Lu, J., & Yu, C.S. (2004). Beyond concern—A privacy-trust—Behavioral intention model of electronic e-commerce. *Information & Management, 42*(1), 127-142.

Milberg, S.J., Smith, H.J., & Burke, S.J. (2000). Information privacy: Corporate management and national regulation. *Organization Science, 11*(1), 35-58.

Miyazaki, A.D., & Fernandez, A. (2001). Consumer perceptions of privacy and security risks for online shopping. *The Journal of Consumer Affairs, 35*(1), 27-55.

O'Brein, D.G., & Yasnoff, W.A. (1999). Privacy, confidentiality, and security in information systems of state health agencies. *American Journal of Preventive Medicine, 16*(4), 351-358.

Olivero, N., & Lunt, P. (2004). Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control. *Journal of Economic Psychology, 25*(2), 243-262.

Ranganathan, C., & Ganapathy, S. (2002). Key dimensions of business-to-consumer Web sites. *Information & Mangement, 39*(6), 457-465.

Smith, H.J. (2001). Information privacy and marketing: What the US should (and shouldn't) learn from Europe. *California Management Review, 43*(2), 8-34.

Strauss, J., & Rogerson, K.S. (2002). Policies for online privacy in the United States and the European Union. *Telematics and Informatics, 19*(2), 173-192.

Xu, Y., Tan, B.C.Y., Hui, K.L., & Tang, W.K. (2003). Consumer trust and online information privacy. *Proceedings of the Twenty-Fourth International Conference on Information Systems*, Seattle, Washington.

KEY TERMS

Cookies: A string of text that a Web browser sends to you while you are visiting a Web page. It is saved on your hard drive, and it saves information about you or your computer. The next time you visit this Web site, the information saved in this cookie is sent back to the Web browser to identify you.

Data Mining: A process by which information is extracted from a database or multiple databases using computer programs to match and merge data and create more information.

Fair Information Practices (FIP): Developed in 1973 by the U.S. Department of Health, Education, and Welfare (HEW) to provide guidelines to protect computerized records. These principles are collection, disclosure, accuracy, security, and secondary use. Some scholars categorize the categories as follows: notice, choice, access, security, and contact information (Culnan, 1999; FTC, 2000).

Opt-In/Opt-Out: A strategy that a business may use to set up a default choice (opt-in) in a form that forces a customer, for example, to accept e-mails or give permission to use personal information, unless the customer deliberately decline this option (opt-out).

Organization for Economic Co-operation and Development (OECD) Guidelines: International guidelines for protecting an individual's privacy (similar to the FIP).

Privacy: The right to be left alone and the right to control and manage information about oneself.

Privacy Seals: A seal that a business may put on its Web site (i.e., Verisign or TRUSTe) to show that it is a trustworthy organization that adheres to its privacy policies.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 475-479, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.10
Enhancing E-Business on the Semantic Web through Automatic Multimedia Representation

Manjeet Rege
Wayne State University, USA

Ming Dong
Wayne State University, USA

Farshad Fotouhi
Wayne State University, USA

ABSTRACT

With the evolution of the next generation Web—the Semantic Web—e-business can be expected to grow into a more collaborative effort in which businesses compete with each other by collaborating to provide the best product to a customer. Electronic collaboration involves data interchange with multimedia data being one of them. Digital multimedia data in various formats have increased tremendously in recent years on the Internet. An automated process that can represent multimedia data in a meaningful way for the Semantic Web is highly desired. In this chapter, we propose an automatic multimedia representation system for the Semantic Web. The proposed system learns a statistical model based on the domain specific training data and performs automatic semantic annotation of multimedia data using eXtensible Markup Language (XML) techniques. We demonstrate the advantage of annotating multimedia data using XML over the traditional keyword based approaches and discuss how it can help e-business.

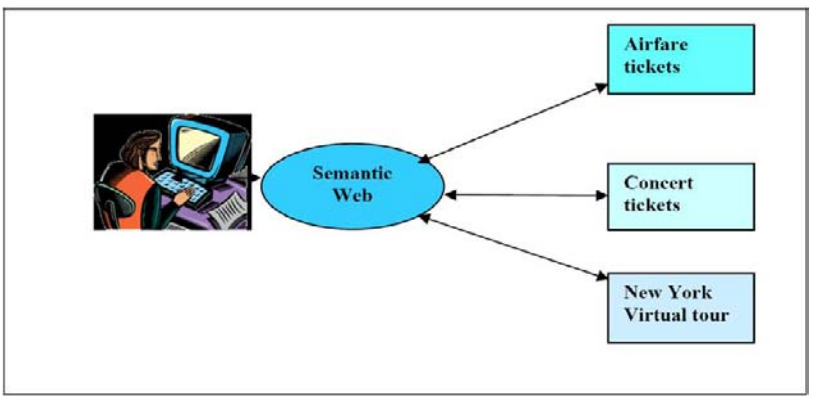
INTRODUCTION

An Internet user typically conducts separate individual e-business transactions to accomplish a certain task. A tourist visiting New York might purchase airfare tickets and tickets to a concert in New York separately. With the evolution of the Semantic Web, as shown in Figure 1, the user can conduct one collaborative e-business transaction for the two purchases. Moreover, he/she can also take a virtual tour of New York city online, which actually might be a collection of all videos, images, and songs on New York appearing anywhere on the World Wide Web. With the continuing growth and reach of the Web, the multimedia data available on it continue to grow on a daily basis. For a successful collaborative e-business, in addition to other kinds of data, it is important to be able to organize and search the multimedia data for the Semantic Web.

With the Semantic Web being the future of the World Wide Web of today, there has to be an efficient way to represent the multimedia data automatically for it. Multimedia data impose a great challenge to document indexing and retrieval as it is highly unstructured and the semantics are implicit in the content of it. Moreover, most of the multimedia contents appearing on the Web have no description available with it in terms of keywords or captions. From the Semantic Web point of view, this information is crucial because it describes the content of multimedia data and would help represent it in a semantically meaningful way. Manual annotation is feasible on a small set of multimedia documents but is not scalable as the number of multimedia documents increases. Hence, performing manual annotation of all Web multimedia data while "moving" them to the Semantic Web domain is an impossible task. This we believe is a major challenge in transforming today's Web multimedia data into tomorrow's Semantic Web data.

In this chapter, we propose a generic automatic multimedia representation solution for the Semantic Web—an XML-based (Bray, Paoli, & Sperberg-McQueen, 1998) automatic multimedia representation system. The proposed system is implemented using images as an example and performs domain-specific annotation using XML. Specifically, our system "learns" from a set of domain-specific training images made available to it a priori. Upon receiving a new image from the Web that belongs to one of the semantic categories the system has learned, the system generates appropriate XML-based annotation for the new image, making it "ready" for the Semantic Web. Although the proposed system has been described from the perspective of images, in general it is

Figure 1. Collaborative e-business scenario on the Semantic Web



applicable to many kinds of multimedia data available on the Web today. To our best knowledge, there has been no work done on automatic multimedia representation for the Semantic Web using the semantics of XML. The proposed system is the first work in this direction.

BACKGROUND

The term *e-business* in general refers to online transactions conducted on the Internet. These are mainly classified into two categories: business-to-consumer (B2C) and business-to-business (B2B). One of the main differences between these two kinds of e-businesses is that B2C, as the name suggests, applies to companies that sell their products or offer services to consumers over the Internet. B2B on the other hand are online transactions conducted between two companies. From its initial introduction in late 1990s, e-business has grown to include services such as car rentals, health services, movie rentals, and online banking. The Web site CIO.com (2006) reports that North American consumers have spent $172 billion shopping online in 2005, up from $38.8 billion in 2000. Moreover, e-business is expected to grow even more in the coming years. By 2010, consumers are expected to spend $329 billion each year online. We expect the evolving Semantic Web to play a significant role in enhancing the way e-business is done today. However, as mentioned in the earlier section, there is a need to represent the multimedia data on the Semantic Web in an efficient way. In the following section, we review some of the related work done on the topic.

Ontology/Schema-Based Approaches

Ontology-based approaches have been frequently used for multimedia annotation and retrieval. Hyvonen, Styrman, and Saarela (2002) proposed ontology-based image retrieval and annotation of graduation ceremony images by creating hierarchical annotation. They used Protégé (n.d.) as the ontology editor for defining the ontology and annotating images. Schreiber, Dubbeldam, Wielemaker, and Wielinga (2001) also performed ontology-based annotation of ape photographs, in which they use the same ontology defining and annotation tool and use Resource Definition Framework (RDF) Schema as the output language. Nagao, Shirai, and Squire (2001) have developed a method for associating external annotations to multimedia data appearing over the Web. Particularly, they discuss video annotation by performing automatic segmentation of video, semiautomatic linking of video segments, and interactive naming of people and objects in video frames. More recently, Rege, Dong, Fotouhi, Siadat, and Zamorano (2005) proposed to annotate human brain images using XML by following the MPEG-7 (Manjunath, 2002) multimedia standard. The advantages of using XML to store meta-information (such as patient name, surgery location, etc.), as well as brain anatomical information, has been demonstrated in a neurosurgical domain. The major drawback of the approaches, mentioned previously, is that the image annotation is performed manually. There is an extra effort needed from the user's side in creating the ontology and performing the detailed annotation. It is highly desirable to have a system that performs automatic semantic annotation of multimedia data on the Internet.

Keyword-Based Annotations

Automatic image annotation using keywords has recently received extensive attention in the research community. Mori, Takahashi, and Oka (1999) developed a co-occurrence model, in which they looked at the co-occurrence of keywords with image regions. Duygulu, Barnard, Freitas, and Forsyth (2002) proposed a method to describe images using a vocabulary of blobs. First, regions are created using a segmentation algorithm. For

each region, features are computed and then blobs are generated by clustering the image features for these regions across images. Finally, a translation model translates the set of blobs of an image to a set of keywords. Jeon, Lavrenko, and Manmatha (2003) introduced a cross-media relevance model that learns the joint distribution of a set of regions and a set of keywords rather than the correspondence between a single region and a single keyword. Feng, Manmatha, and Lavrenko (2004) proposed a method of automatic annotation by partitioning each image into a set of rectangular regions. The joint distribution of the keyword annotations and low-level features is computed from the training set and used to annotate testing images. High annotation accuracy has been reported. The readers are referred to Barnard, Duygulu, Freitas, and Forsyth (2003) for a comprehensive review on this topic. As we point out in the section, "XML-Based Annotation," keyword annotations do not fully express the semantic meaning embedded in the multimedia data. In this paper, we propose an Automatic Multimedia Representation System for the Semantic Web using the semantics of XML, which enables efficient multimedia annotation and retrieval based on the domain knowledge. The proposed work is the first attempt in this direction.

PROPOSED FRAMEWORK

In order to represent multimedia data for the Semantic Web, we propose to perform automatic multimedia annotation using XML techniques. Though the proposed framework is applicable to multimedia data in general, we provide details about the framework using image annotations as a case study.

XML-Based Annotation

Annotations are domain-specific semantic information assigned with the help of a domain expert to semantically enrich the data. The traditional approach practiced by image repository librarians is to annotate each image manually with keywords or captions and then search on those captions or keywords using a conventional text search engine. The rationale here is that the keywords capture the semantic content of the image and help in retrieving the images. This technique is also used by television news organizations to retrieve file footage from their videos. Such techniques allow text queries and are successful in finding the relevant pictures. The main disadvantage with manual annotations is the cost and difficulty of scaling it to large numbers of images.

MPEG-7 (Manjunath, 2002, p. 8) describes the content—"the bits about the bits"—of a multimedia file such as an image or a video clip. The MPEG-7 standard has been developed after many rounds of careful discussion. It is expected that this standard would be used in searching and retrieving for all types of media objects. It proposes to store low-level image features, annotations, and other meta-information in one XML file that contains a reference to the location of the corresponding image file. XML has brought great features and promising prospects to the future of the Semantic Web and will continue to play an important role in its development. XML keeps content, structure, and representation apart and is a much more adequate means for knowledge representation. It can represent semantic properties through its syntactic structure, that is, by the nesting or sequentially ordering relationship among elements (XML tags). The advantage of annotating multimedia using XML can best be explained with the help of an example. Suppose we have a New York image (shown in Figure 2) with keywords annotation of Statue of Liberty, Sea, Clouds, Sky. Instead of simply using keywords as annotation for this image, consider now that the same image is represented in an XML format.

Note that the XML representation of the image can conform to any domain-specific XML schema. For the sake of illustration, consider

the XML schema and the corresponding XML representation of the image shown in Figure 3. This XML schema stores foreground and background object information along with other meta-information with keywords along various paths of the XML file. Compared with keyword-based approaches, the XML paths from the root node to the keywords are able to fully express the semantic meaning of the multimedia data. In the case of the New York image, semantically meaningful XML annotations would be "image/semantic/foreground/object=Statue of Liberty, image/semantic/foreground/ object = Sea, image/semantic/ background/ object = Sky, image/semantic/background /object =Clouds". The semantics in XML paths provides us with an added advantage by differentiating the objects in the foreground and background and giving more meaningful annotation.

We emphasize that the annotation performed using our approach is domain-specific knowledge. The same image can have different annotation under a different XML schema that highlights certain semantic characteristics of importance pertaining to that domain knowledge. We simply use the schema of Figure 3 that presents image foreground and background object information as a running example.

Overview of System Architecture

The goal of the proposed system is to represent multimedia data obtained from the Web in a meaningful XML format. Consequently, this data can be "moved" to the Semantic Web in an automatic and efficient way. For example, as shown in Figure 4, the system first receives an image from the Web. The image could be received by a *Web image provider* which is an independent module outside of the system that simply fetches domain-specific images from the Web and passes them onto our system. The Web image provider could also be a "Web spider" that "crawls" among domain-specific Web data sources and procures relevant images. The image is then preprocessed by two other modules, namely, *image divider* and *feature extractor*. An image usually contains several regions. Extracting low-level features from different image regions is typically the first step of automatic image annotation since regions may have different contents and represent different semantic meaning. The image regions could be determined through either image segmentation (Shi & Malik, 1997) or image cutting in the image divider. For low-level feature extraction, we used some of the features standardized by MPEG-7.

The low-level features extracted from all the regions are passed on to the *automatic annotator*. This module learns a statistical model that links image regions and XML annotation paths from a set of domain-specific training images. The training image database can contain images belonging to various semantic categories represented and annotated in XML format. The annotator learns to annotate new images that belong to at least one of the many semantic categories that the

Figure 2. Comparison of keyword annotation and XML-path-based annotation

Image	Original Annotation	XML annotation
	Statue of Liberty Sea Sky Clouds	image/semantic/foreground/object = Statue of Liberty image/semantic/foreground/object = Sea image/semantic/background/object= Sky image/semantic/background/object= Clouds

Figure 3. An example of an XML schema and the corresponding XML representation of an image

```
<!DOCTYPE image[
<!ELEMENT image(semantic, meta, features) >
<!ELEMENT semantic(category, background, foreground) >
<!ELEMENT meta(id, caption, imageOwner, dateOfCreation) >
<!ELEMENT features(color, texture ) >
<!ELEMENT background (object+)>
<!ELEMENT foreground(object+)>
<!ELEMENT color( #PCDATA )>
<!ELEMENT category(#PCDATA) >
<!ELEMENT object(#PCDATA) >
<!ELEMENT id(#PCDATA) >
<!ELEMENT caption(#PCDATA) >
<!ELEMENT imageOwner(#PCDATA) >
<!ELEMENT dateOfCr eation(#PCDATA) >
<!ELEMENT texture (#PCDATA) >
]>
```

```
<image>
  <semantic>
    <category>New York Images</category>

    <background>
      <object>sky</object>
      <object>clouds</object>
    </back ground>

    <foreground>
      <object>statue of liberty</object>
      <object>sea</object>
    </foreground>
  </semantic>

  <meta>
    <id>396003</id>
          <caption>New York</caption>
          <imageOwner>New York
                Galleries
          </imageOwner>
          <dateOfCreation>04/10/2005</dateOfCreation>
  </meta>

  <features>
          <color> 0.65227 0.36126 0.62181…….</color>
          < texture> 0.33782 0.30867 0.31427..</texture >
  </features>
</image>
```

annotator has been trained on. The output of the automatic annotator is an XML representation of the image.

Statistical Model for Automatic Annotation

In general, image segmentation is a computationally expensive as well as an erroneous task (Feng et al., 2004). As an alternative simple solution, we have the image divider partition each image into a set of rectangular regions of equal sizes. The feature extractor extracts low-level features from each rectangular region of every image and constructs a feature vector. By learning the joint probability distribution of XML annotation paths and low-level image features, we perform the automatic annotation of a new image.

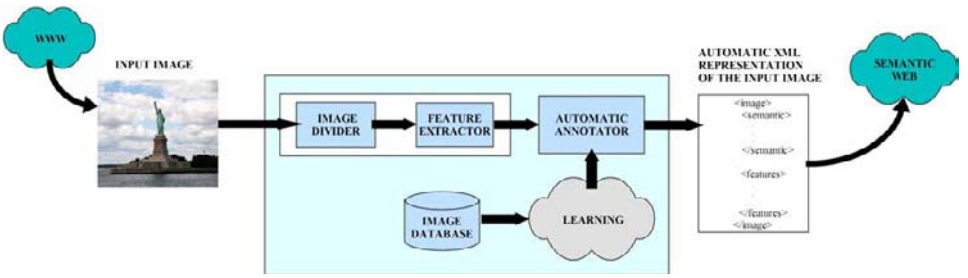
Let X denote the set of XML annotation paths, T denote the domain-specific training images in XML format, and let t be an image belonging to T. Let x_t be a subset of X containing the annotation paths for t. Also, assume that each image is divided into n rectangular regions of equal size.

Consider a new image q not in the training set. Let $f_q = \{f_{q1}, f_{q2}, \ldots..f_{qn}\}$ denote the feature vector for q. In order to perform automatic annotation of q, we model the joint probability of f_q and any arbitrary annotation path subset x of X as follows,

$$P(x, f_q) = P(x, f_{q1}, f_{q2}, \ldots..., f_{qn}) \quad (1)$$

We use the training set T of annotated images to estimate the joint probability of observing *x* and $\{f_{q1}, f_{q2,\ldots\ldots..,} f_{qn}\}$ by computing the expectation over

Figure 4. System architecture



all the images in the training set.

(2)

$$P(x, f_{q1}, f_{q2}, \ldots, f_{qn}) = \sum_{t \in T} P(t)\, P(x, f_{q1}, f_{q2}, \ldots, f_{qn} \mid t)$$

We assume that the events of observing x and $f_{q1,}$ $f_{q2,}$ f_{qn} are mutually independent of each other and express the joint probability in terms of P_A, P_B and P_C as follows:

(3)

$$P(x, f_{q1}, f_{q2}, \ldots f_{qn}) = \sum_{t \in T} \{ P_A(t) \prod_{a} P_B(f_a|t) \prod_{path \in x} P_C(path|t) \prod_{path \notin x} (1 - P_C(path|t)) \}$$

where P_A is the prior probability of selecting each training image, P_B is the density function responsible for modeling the feature vectors, and P_C is a multiple Bernoulli distribution for modeling the XML annotation paths.

In the absence of any prior knowledge of the training set, we assume that P_A follows a uniform prior and can be expressed as:

$$P_A = \frac{1}{\| T \|} \quad (4)$$

where ||T|| is the size of the training set. For the distribution P_B, we use a nonparametric, kernel-based density estimate:

$$P_B(f|t) = \frac{1}{n} \sum_i \frac{\exp\{-(f - f_i)^T \Sigma^{-1} (f - f_i)\}}{\sqrt{2^k \prod^k |\Sigma|}} \quad (5)$$

where f_i belongs to $\{f_1, f_2, \ldots, f_n\}$ the set of all low-level features computed for each rectangular region of $\sum$ image t. is the diagonal covariance matrix which is constructed empirically for best annotation performance.

In the XML representation of images, every annotation path can either occur or might not occur at all for an image. Moreover, as we annotate images based on object presence and not on prominence in an image, an annotation path if it occurs can occur—at most—once in the XML representation of the image. As a result, it is reasonable to assume that the density function P_C follows a multiple Bernoulli distribution as follows:

$$P_C(path|t) = \frac{(\gamma \alpha_{path,t} + N_{path})}{(\gamma + \| T \|)} \quad (6)$$

where γ is a smoothing parameter, if the path occurs in the annotation of image *t*, else it is zero. N_{path} is the total number of training images that contain this *path* in their annotation.

EXPERIMENTAL RESULTS

Our image database contains 1,500 images obtained from the Corel data set, comprising 15 image categories with 100 images in each category. The Corel image data set contains images from different semantic categories with keyword annotations performed by Corel employees. In order to conduct our experiments, we require a

training image database representing images in XML format. Each XML file should contain annotation, low-level features, and other meta-information stored along different XML paths. In the absence of such a publicly available data, we had to manually convert each image in the database to an XML format conforming to the schema shown in Figure 3. We performed our experiments on five randomly selected image categories. Each image category represents a distinct semantic concept. In the experiments, 70% of the data are randomly selected as the training set while the remaining were used for testing.

Automatic Annotation Results

Given a test image, we calculate the joint probability of the low-level feature vector and the XML annotation paths in the training set. We select the top four paths with the highest joint probability as the annotation for the image. Compared with other approaches in image annotation (Duygulu et al., 2002; Feng et al., 2004), our annotation results provide more meaningful description of a given image.

Figure 5 shows some examples of our annotation results. We can clearly see that the XML-path-based annotation contains richer semantic meaning than the original keyword provided by Corel. We evaluate the image annotation performance in terms of recall and precision. The recall and precision for every annotation path in the test set is computed as follows:

$$\text{recall} = \frac{q}{r}$$

$$\text{precision} = \frac{q}{s}$$

where q is the number of images correctly annotated by an annotation path, r is the number of images having that annotation path in the test set, and s is the number of images annotated by the same path. In Table 1 we report the results for all the 148 paths in the test set as well as the 23 best paths as in Duygulu et al. (2002) and Feng et al. (2004).

Retrieval Results

Given specific query criteria, XML representation helps in efficient retrieval of images over

Figure 5. Examples of top annotation in comparison with Corel keyword annotation

Image		
Corel keyword Annotation	plane, jet, wheels, sky	sky, clouds, train, tracks
Automatic XML based Annotation	image/semantic/background/object=sky, image/semantic/foreground/object=plane, image/semantic/foreground/object=jet, image/semantic/foreground/object=wheels	image/semantic/background/object=sky, image/semantic/background/object=clouds, image/semantic/forgeround/object=train, image/semantic/forgeround/object=tracks,

Table 1. Annotation results

Number of Paths with recall > 0 is 50		
Annotation Results	**Results on all 148 paths**	**Results on top 23 paths**
Mean per-path recall	0.22	0.83
Mean per-path precision	0.21	0.73

the Semantic Web. Suppose a user wants to find images that have an *airplane* in the background and *people* in the foreground. State-of-the-art search engines require the user to supply individual keywords such as "airplane," "people," and so forth or any combination of keywords as a query. The union of the retrieved images of all possible combinations of the aforementioned query keywords is sure to have images satisfying the user specified criteria.

However, a typical search engine user searching for images is unlikely to view beyond the first 15-20 retrievals, which may be irrelevant in this case. As a result, the user query in this scenario is unanswered in spite of images satisfying the specified criteria being present on the Web. With the proposed framework, the query could be answered in an efficient way. Since all the images on the Semantic Web are represented in an XML format, we can use XML querying technologies such as XQuery (Chamberlin, Florescu, Robie, Simeon, & Stefanascu, 2001) and XPath (Clark & DeRose, 1999) to retrieve images for the query "image/semantic/background/object = plane & image/semantic/foreground/object = people". This is unachievable with keyword-based queries and hence is a major contribution of the proposed work.

Figure 6 shows some examples of the retrieval results. In Table 2, we also report the mean average precision obtained for ranked retrieval as in Feng et al. (2004).

Since the proposed work is the first one of its kind to automatically annotate images using XML

Figure 6. Ranked retrieval for the query image/semantic/background/object = "sky"

Table 2. Mean average precision results

All 148 paths	**Paths with recall > 0**
0.34	0.38

paths, we were unable to make a direct comparison with any other annotation model. However, our annotation and retrieval results are comparable to the ones obtained by Duygulu et al. (2002) and Feng et al. (2004).

CONCLUSION AND DISCUSSION

With the rapid development of digital photography, more and more people are able to share their personal photographs and home videos on the Internet. Many organizations have large image and video collections in digital format available for online access. For example, film producers advertise movies through interactive preview clips. News broadcasting corporations post photographs and video clips of current events on their respective Web sites. Music companies have audio files of their music albums made available to the public online. Companies concerning the travel and tourism industry have extensive digital archives of popular tourist attractions on their Web sites. As this multimedia data is available—although scattered across the Web—an efficient use of the data resource is not being made. With the evolution of the Semantic Web, there is an immediate need for a semantic representation of these multimedia resources. Since the Web is an infinite source of multimedia data, a manual representation of the data for the Semantic Web is virtually impossible. We present the Automatic Multimedia Representation System that annotates multimedia data on the Web using state-of-the art XML technologies, thus making it "ready" for the Semantic Web.

We show that the proposed XML annotation has a more semantic meaning over the traditional keyword-based annotation. We explain the proposed work by performing a case study of images, which in general is applicable to multimedia data available on the Web.

The major contributions of the proposed work from the perspective of multimedia data sources representation can be stated as follows:

- **Multimedia annotation:** Most of the multimedia data appearing on the World Wide Web are unannotated. With the proposed system, it would be possible to annotate this data and represent it in a meaningful XML format. This we believe would enormously help in "moving" multimedia data from World Wide Web to the Semantic Web.
- **Multimedia retrieval:** Due to representation of multimedia data in XML format, the user has an advantage to perform a complex semantic query instead of the traditional keyword based.
- **Multimedia knowledge discovery:** By having multimedia data appear in an XML format, it will greatly help intelligent Web agents to perform Semantic Web mining for multimedia knowledge discovery.

From an e-business point of view, semantically represented and well-organized Web data sources can significantly help the future of a *collaborative* e-business by the aid of intelligent Web agents. For example, an agent can perform autonomous tasks such as interact with travel Web sites and obtain attractive vacation packages where the users can bid for a particular vacation package or receive the best price for a book across all the booksellers. It is important to note that in addition to multimedia data, once other data sources are also represented in accordance with the spirit of the Semantic Web, the opportunities for collaborative e-business tasks are endless.

REFERENCES

Barnard, K., Duygulu, P., Fretias, N., Forsyth, D., Blei, D., & Jordan, M. I. (2003). Matching words and pictures. *Journal of Machine Learning Research, 3,* 1107-1135.

Bray, T., Paoli, J., & Sperberg-McQueen, C. M. (1998, February 10). *Extensible markup language (XML) 1.0*. Retrieved October 15, 2006, from http://www.w3.org/TR/1998/REC-xml-19980210

Chamberlin, D., Florescu, D., Robie, J., Simeon, J., & Stefanascu, M. (2001). *XQuery: A query language for XML*. Retrieved from http://www.w3.org/TR/xquery

CIO.com. (2006). *The ABCs of e-commerce*. Retrieved October 15, 2006, from http://www.cio.com/ec/edit/b2cabc.html

Clark, J., & DeRose, S. (1999, November 16). *XML path language (XPath) Version 1.0*. Retrieved August 31, 2006, from http://www.w3.org/TR/xpath

Duygulu, P., Barnard, K., Freitas, N., & Forsyth, D. (2002). Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of European Conference on Computer Vision, 2002* (LNCS 2353, pp. 97-112). Berlin; Heidelberg: Springer.

Feng, S. L., Manmatha, R., & Lavrenko, V. (2004). Multiple Bernoulli relevance models for image and video annotation. In *Proceedings of IEEE Conference on Computer Vision Pattern Recognition, 2004* (Vol. 2, pp. 1002-1009).

Hyvonen, E., Styrman, A., & Saarela, S. (2002). Ontology-based image retrieval. In *Towards the Semantic Web and Web services, Proceedings of XML Finland Conference*, Helsinki, Finland (pp. 15-27).

Jeon, J., Lavrenko, V., & Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada (pp. 119-126). New York: ACM Press.

Manjunath, B. S. (2002). *Introduction to MPEG-7: Multimedia content description interface*. John Wiley and Sons.

Mori, Y., Takahashi, H., & Oka, R. (1999). Image-to-word transformation based on dividing and vector quantizing images with words. In *Proceedings of First International Workshop on Multimedia Intelligent Storage and Retrieval Management.*

Nagao, K., Shirai, Y., & Squire, K. (2001). Semantic annotation and transcoding: Making Web content more accessible. *IEEE Multimedia Magazine, 8*(2), 69-81.

Protégé. (n.d.). (Version 3.1.1) [Computer software]. Retrieved February 19, 2006, from http://protege.stanford.edu/index.html

Rege, M., Dong, M., Fotouhi, F., Siadat, M., & Zamorano, L. (2005). Using Mpeg-7 to build a human brain image database for image-guided neurosurgery. In *Proceedings of SPIE International Symposium on Medical Imaging*, San Diego, CA (Vol. 5744, pp. 512-519).

Schreiber, A. T., Dubbeldam, B., Wielemaker, J., & Wielinga, B. (2001). Ontology based photo annotation. *IEEE Intelligent Systems, 16*(3), 66-74.

Shi, J., & Malik, J. (1997). Normalized cuts and image segmentation. In *Proceedings of 1997 IEEE Conference on Computer Vision Pattern Recognition*, San Juan (pp. 731-737).

This work was previously published in Semantic Web Technologies and E-Business: Toward the Integrated Virtual Organization and Business Process Automation, edited by A. Salam and J. Stevens, pp. 154-168, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.11
Digital Multimedia Broadcasting (DMB) in Korea:
Convergence and its Regulatory Implications

Seung Baek
Hanyang University, Korea

Bong Jun Kim
Korea Telecommunications (KT) Marketing Laboratory, Korea

ABSTRACT

The launch of portable Internet, alongside mobile Internet technology and cellular technology, is a new milestone, converging wireless with wired technology. Along with these new technologies, a new telecommunication service has been introduced and has received much attention from the Korean public. This is the Digital Multimedia Broadcasting (DMB) service. DMB is a digital multimedia service combining telecommunications with broadcasting technologies. DMB enables users to watch various multimedia contents on their phone screens while they are on the move. Since DMB services in Korea are the first in the world, the Korean Government has much interest in DMB services. However, the repeated failures in establishing a regulatory framework for DMB and ill-defined roles of players in the DMB industry interfere the diffusion of DMB in the Korean market. As the convergence of broadcasting and telecommunications makes progress, proper modifications of existing regulatory frameworks should be made in order to guarantee success of DMB service in Korea. This chapter reviews DMB technology, its business model, its market structure, and its policy. In particular, it explores business opportunities around DMB services and identifies major issues that must be solved to launch DMB services successfully.

INTRODUCTION

In Korea, the number of Internet users has been growing rapidly, nearly doubling each year since

1997. What is even more interesting is that most Internet users subscribe to high-speed Internet service. In 2001, the number of subscribers per 100 people was 21.8 people in Korea (about 40% of all Internet users), 4.5 people in the United States (about 9% of Internet users), and 2.2 people in Japan (5% of Internet users). This dramatic expansion of the high-speed Internet service has even received worldwide attention. The International Telecommunication Union (ITU) and the Organisation for Economic Co-operation and Development (OECD) announced that Korea ranked first in the diffusion of high-speed Internet service. Ninety-seven percent of all households in Korea have some way of connecting to the Internet and 60% of all households in Korea access the high-speed Internet.

In Korea, the phenomenal growth of ownership of cellular phones was not a government initiative, rather a private industry-driven one. Due to the highly efficient electronics industry which was able to manufacture low-cost/high-capacity cellular phones, the Korean public quickly adopted the use of cellular phones in their everyday lives. According to statistics, almost 90% of all adults now own a cellular phone, which makes Korea the country with highest ownership of cellular phones in the world. Recently, the use of mobile Internet through various handsets, such as cellular phones and personal digital assistants (PDAs), has become popular.

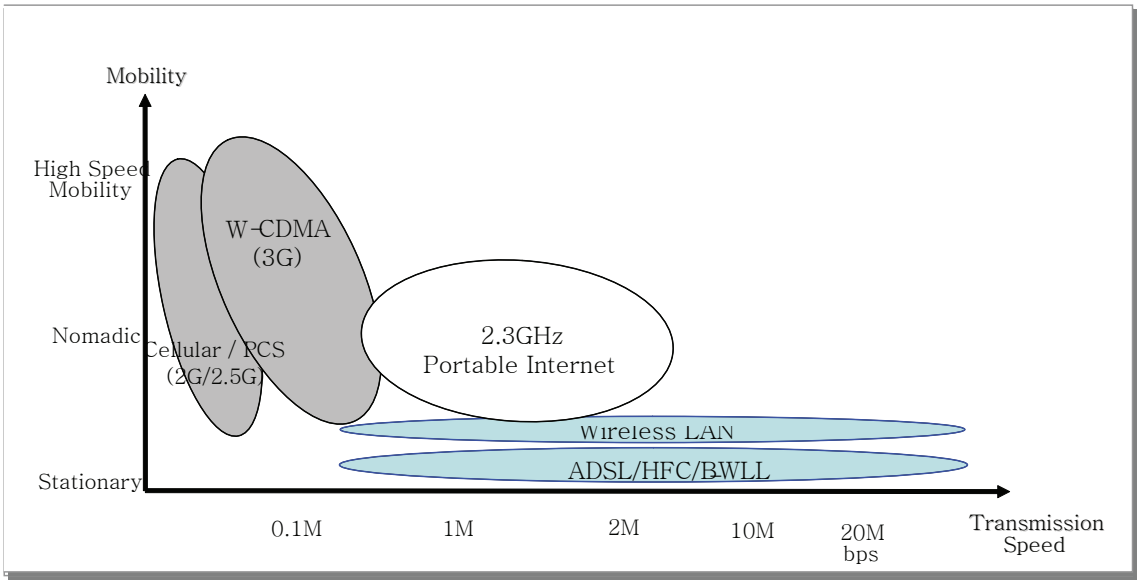
Now, many Korean users have utilized the Internet for personal communications (e.g., e-mail) and information searching. As the user population of the high-speed Internet service is growing quickly in Korea, many users are more inclined to use the Internet for multimedia entertainment, such as games, movies, and music. The high-speed Internet service shifts its main usage to entertainment. In terms of mobile Internet, its main usage is also concentrated on entertainment, such as ring/avatar downloads.

The launch of portable Internet, alongside mobile Internet technology and cellular technology, is a new milestone, converging wireless with wired technology. Along with these new technologies, a new telecommunication service has been introduced and has received much attention from the Korean public. This is the Digital Multimedia Broadcasting (DMB) service. DMB is a digital multimedia service combining telecommunications with broadcasting technologies. DMB enables users to watch various multimedia contents on their phone screens while they are on the move. By combining telecommunication and broadcasting technologies, DMB adds tremendous value to broadcasting. Traditionally, broadcasting technology is used to transmit information to many unspecified persons (one-to-many, one-way communication), and it is very difficult to watch TV while on the move. Whereas telecommunication technology allows individual communications (one-to-one, two-way communication) and it is easy to provide personalized services.

DMB, selected as one of the 10 new-growth engine sectors—in other words, one of the 10 most promising industries to propel Korea toward its goal of passing the $20,000 mark in GDP per capita—opens up a vast new horizon for broadcasting, making the most of strengths and specificities of different media, including terrestrial, cable, and satellite broadcasts. Economic and social ripple effects to be expected from DMB also are certainly substantial.

The question remains, however, as to whether DMB can indeed stake out its own market in Korea, as CDMA or the high-speed Internet has done in the past. The answers to this question are so far mutually contradictory even among experts. Figure 1 illustrates the respective positions of different types of telecommunication services in Korea, from which the place of DMB can be roughly deduced. More optimistic onlookers hold the view that DMB, by overcoming the one-way services thus far provided by the mobile Internet or the HSDPA-based mobile Internet, and by wooing over customers with more competitive prices, will be able to create its own niche. The

Figure 1. Telecommunication services positioning in Korea



opposing view holds that the market strategy of putting faith in DMB as a killer application is blunted by the fact that many of its supposedly killer service features are already offered by mobile Internet or HSDPA-based mobile Internet services. This redundancy has the implication that consumers constituting the demand source for DMB may also overlap with those making up the existing portable multimedia service market. In other words, DMB may turn out to be merely a complementary service, remedying some of the weaknesses of competitor services.

As predictions on the market viability of DMB remain divided among onlookers, what place, then, will DMB actually occupy within Korea's telecommunications market and how will it evolve within it? And what would be the response strategies by communications carriers? The objective of this chapter is to offer answers to some of these essential questions regarding the prospect of DMB in the domestic market.

DIGITAL MULTIMEDIA BROADCASTING (DMB)

DMB Technology

Broadcasting is quickly moving into the era of digitalization by replacing traditional analog broadcasting technology with digital broadcasting technology. In 1997, the Ministry of Information and Communications (MIC) created a committee for terrestrial digital broadcasting. This committee mainly focused on the shift from analog radio broadcasting to digital radio broadcasting. At that time, the committee investigated various ways to launch the terrestrial Digital Audio Broadcasting (DAB) service into the Korean market. As telecommunication companies expressed their interests in DAB, many companies became interested on satellite DAB as well as terrestrial DAB. DAB services enable customers to receive CD-like quality radio programs, even in the car, without any annoying interference and signal distortion. Aside from distortion-free reception and CD-quality sound, DAB offers further advantages as it has been designed for the multimedia age. DAB can

carry not only audio, but also text, pictures, data, and even videos. In 2002, by adding multimedia components to DAB, MIC announced a new service, called DMB (digital multimedia broadcasting), as a way to accelerate the convergence of telecommunication and broadcasting services. DMB services can be categorized into terrestrial DMB and satellite DMB.

Terrestrial DMB

Transmission of TV and radio signals using ground-based transmitters in a terrestrial network is traditionally the most used and known distribution form. End users can receive the signals through reception roofs or in-house antennas. This distribution form can also be used for digital TV. As shown in Figure 2, once terrestrial DMB integrates various multimedia contents by using multiplexers and orthogonal frequency division multiplex (OFDM) modulators, it transmits these integrated multimedia contents through ground-based transmitters. In 2001, MIC chose Eureka-147 as a standard for terrestrial DMB.

Satellite DMB

Unlike terrestrial DMB, satellite DMB transmits signals through satellites. Terrestrial DMB normally covers limited areas, whereas satellite DMB can transmit signals far away from a country's border. In addition, satellite DMB enables the signals to be transmitted with high technical quality. The high capacity of satellite DMB provides the opportunities to broadcast more channels and various digital services. However, compared with terrestrial DMB, satellite DMB is not cost efficient. As shown in Figure 3, satellite DMB transmits signals through satellites. For fringe areas, it retransmits the signals by using gap fillers. This is a major difference between regular satellite broadcasting and satellite DMB. The world's first DMB satellite, Hanbyol (MBSat), was successfully launched in March 2004. The

Figure 2. Terrestrial DMB

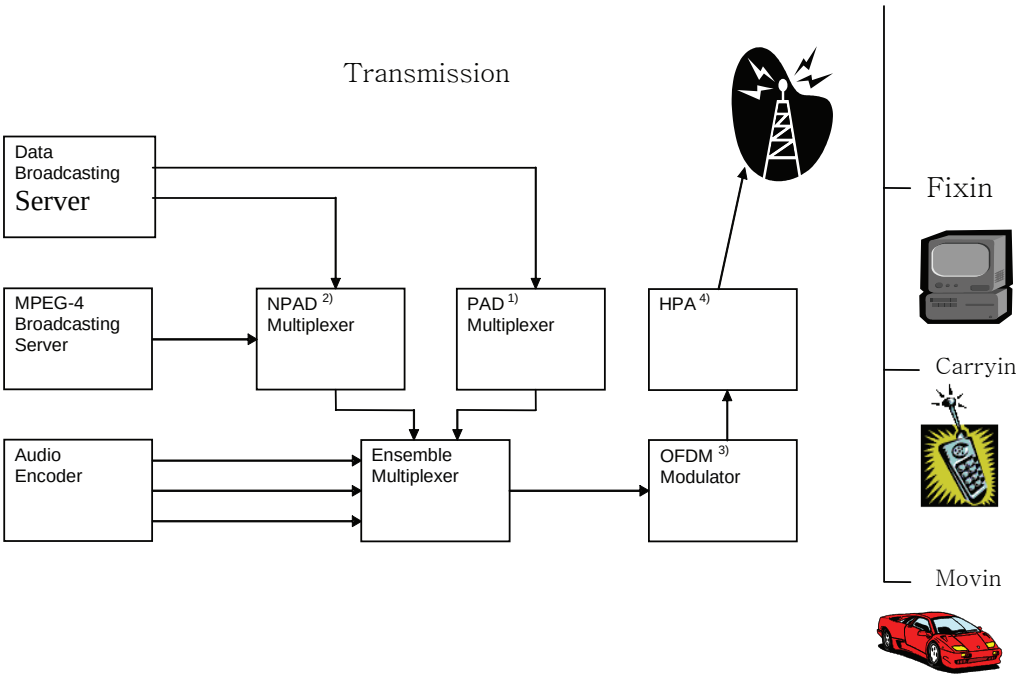
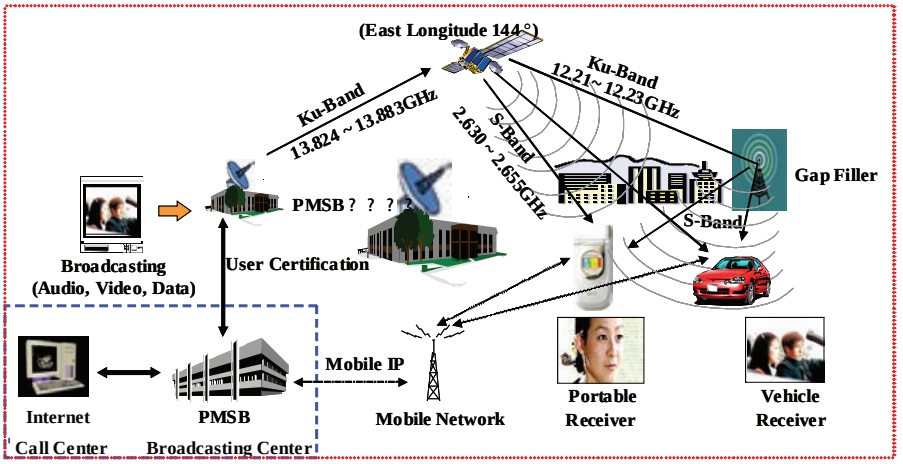


Figure 3. Satellite DMB



satellite was launched by a joint venture between SK Telecom affiliated TU Media and Japan's Mobile Broadcasting Corporation (MBCo). SK Telecom and MBCo are responsible for 34.66% and 65.34% of the total cost for launching the satellite, respectively. Besides SK Telecom, KT, a giant Korean telecommunication company, also expressed its interests in satellite DMB and jumped into the DMB business. From ITU (International Telecommunication Union), KT already occupied frequency band that it is going to use for its satellite DMB service in future. Since MIC announced DMB services in 2002, two giant Korean telecommunication companies, SK Telecom and KT, competed with each other to gain a competitive advantage in the evolving satellite DMB industry. Since two companies plan to use different standards for their own DMB services (SK Telecom uses System E and KT uses System A), their competitions were fierce, rather than collaborative. In 2003, MIC selected System E as a standard for satellite DMB. At that moment, by focusing more on wireless Internet business, KT gave up the satellite DMB business, and SK Telecom-affiliated TU Media has led the satellite DMB business ever since.

As the voice communication market has been saturated, telecommunication companies and MIC have looked for ways to make profit by using existing infrastructures. Recently, MIC announced the IT 8-3-9 Strategy. The IT 8-3-9 Strategy proposes eight new telecommunication services, three telecommunication infrastructures, and nine application areas as concentrated items that the Korean Government has decided to foster (see Table 1). It is an ambitious plan by the Korean Government to promote its competitiveness in the global market. Since DMB services in Korea are the first in the world, the Korean Government has much interest in DMB services. On the other hand, it imposes new political and regulatory challenges and makes rethinking and redesigns of the existing regulatory framework.

DMB Services and Business Models

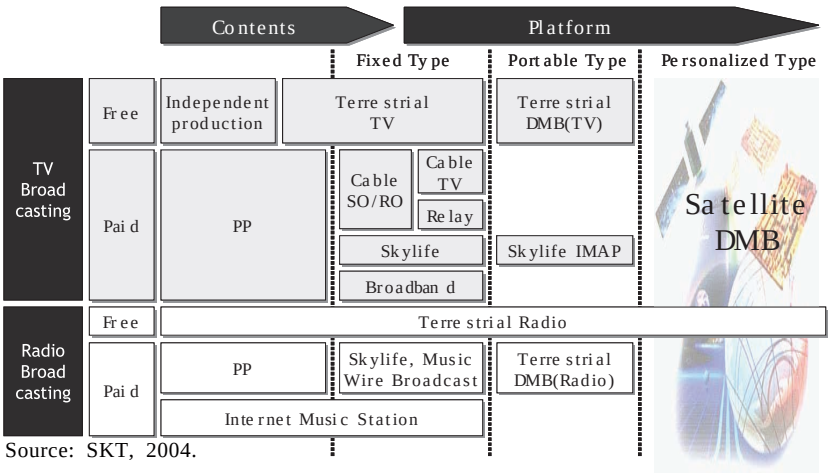
DMB can be defined as a digital broadcasting medium, capable of providing stable transmission of CD-quality audio, data, and video content to fixed or portable reception devices. It is divided into two categories depending on the signals used: terrestrial DMB and satellite DMB. Figure 4 shows the differences between existing broadcasting services and DMB services. Both terrestrial DMB and satellite DMB allow customers to watch TV on their various handsets, such

Table 1. IT 8-3-9 strategy

Telecommunication Services	Infrastructures	Applications
• WiBro Service • DMB Service • Home Network Service • Telemetries Service • RFID Service • W_CDMA Service • Terrestrial DTV • VoIP	• BCN • U-Sensor Network • Ipv6	• Next Generation Mobile Telecommunications • Digital TV • Home Network • IT SoC • Next Generation PC • Embedded • Digital Contents • Telemetries • Intelligent Robot

Source: MIC, 2004.

Figure 4. DMB service areas



Source: SKT, 2004.

as PDAs or cellular phones, while on the move. Furthermore, by integrating existing telecommunication infrastructures and the Internet, DMB makes interactive, and personalized broadcasting possible.

In principle, terrestrial DBM is a free service for customers. Major income sources of terrestrial DMB operators heavily depend on advertisement sales, whereas satellite DMB operators generate income based on subscription charges. Therefore, it is very critical for satellite DMB to differentiate its contents from terrestrial DMB and to provide its services for a reasonable price. On the other hand, it is crucial for terrestrial DMB to acquire various contents in order to compete with satellite DMB.

The DMB service of Korea is one of the national new-growth powers that can provide domestic telecommunication and broadcasting markets with vital energies. Furthermore, DMB, as a national exporting industry, is expected to have strong ripple effects on the Korean economy.

Table 2. DMB service economic spread effect

	2005	2006	2007	2008	2009	2010
Production Effect (Billion Won)	5,924 1	1,995	20,496	27,886	36,917	43,660
Employment Effect (Thousand People)	6.8	13.7	23.1	31.1	40.9	48.0

Figure 5. Equity composition of TU Media

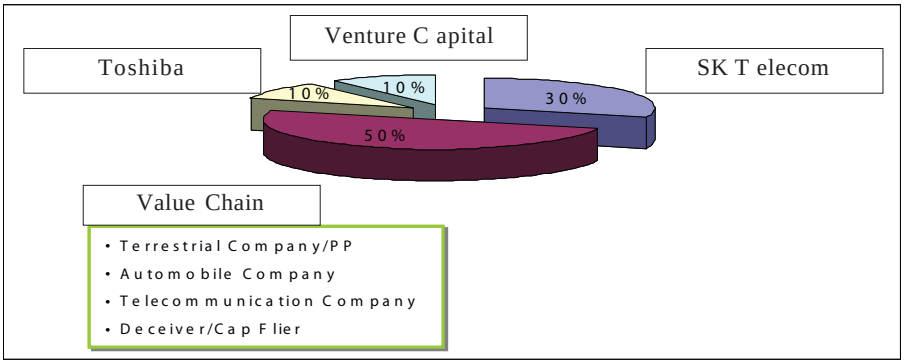


Table 2 summarizes the economic effects of DMB service.

The industry of terrestrial DMB is supposed to be led by three major Korean terrestrial broadcasting companies (KBS, MBC, and SBS). Compared with the satellite DMB industry led by SK Telecom, terrestrial DMB industry does not have many business leaders who are able to or intend to invest large amount of money into terrestrial DMB. From all aspects, including service variety, service coverage, and service quality, terrestrial DMB is inferior to satellite DMB. However, to diffuse DMB service into Korean market at the beginning, the role of terrestrial DMB is crucial due to free service. Since many companies, especially broadcasting companies, cannot guarantee the returns to their investments, they hesitate to participate in terrestrial DMB business actively. At this time, KBS, the nation's representative public broadcaster, is preparing for a terrestrial DMB service, from developing technology of terrestrial DMB to experimenting media. Since KBS has a monopoly in the terrestrial DMB market, other competing broadcasting companies, such as MBC and SBS, recently decided to invest quota in satellite DMB which can be a competitor of terrestrial DMB. It causes conflicts among the three major terrestrial broadcasting companies. From a public broadcaster's point of view, KBS cannot participate in launching charged services like other private broadcasters.

In terms of satellite DMB, SK Telecom, a market leader of the mobile communication, and KT, a market leader of wire communication, were preparing for satellite DMB enterprise together. However, the recent withdrawal of KT from satellite DMB enterprise enables SKT-affiliated TU Media to prepare a satellite DMB enterprise by itself (see Figure 5).

MARKET ISSUES

The Relationship Between Terrestrial DMB and Satellite DMB Players

With the exception that terrestrial DMB and satellite DMB do not share the same frequency band and use different transmission media/paths, they are virtually identical services. Regarding to revenue sources, the main income source of terrestrial DMB is advertising, whereas the income source of satellite DMB is a subscription charge. In terms of coverage, unlike satellite DMB providing nationwide services from the beginning, terrestrial DMB service is restricted to the Seoul and several metropolitan areas early in its commercialization. As DMB service becomes popular, it will offer national coverage in conjunction with DTV migration. Considering the traditional reticence exhibited by the Korean public toward fee-based services, without retransmission by terrestrial TVs of satellite DMB, satellite DMB service may find itself at a significant competitive disadvantage.

With new developments taking place in the telecommunication and broadcasting markets, it might be difficult to clearly define roles of terrestrial DMB and satellite DMB and their mutual relationship. If policy assigns and regulates the respective territories of these two services, it may help to avoid the overlapping of these two services, but it may also prove to be an overintervention, hurting DMB market takeoff. On the other hand, by entirely depending on the market and by lifting or loosening restrictions, the two services are likely to develop a competing relationship with each other. It is quite a difficult task for the Korean Government to balance between restrictions and free competition.

Another potential problem is passive investment of existing terrestrial broadcasting companies into terrestrial DMB. Unlike satellite DMB being expected continuous large-scale investments by major telecommunication companies and others, terrestrial DMB might experience difficulty in obtaining such large investments, due to the lack of well-defined business models. If such scenarios turn out to be true, this will certainly be a hindrance to continuous content development. It can ultimately result in the stagnation of the terrestrial DMB market. Although the fact that terrestrial DMB offers free services might be an appealing formula for the market in the short run, terrestrial DMB might be beaten out of the competition with satellite DMB, owing to the lack of resources for developmental activities. Moreover, to use network effects, satellite DMB players may reduce subscription prices in time. In the beginning, the Korean government might need to provide strong incentives that enable terrestrial broadcasting companies to invest huge amounts of money into terrestrial DMB.

From the viewpoints of service coverage, diversity of contents, and investment size and intention, things are not going to get any brighter for terrestrial DMB. Because the Korean government has a tendency to implement a market-centered policy in the DMB market, the relationship between the two services may inevitably turn out to be competitive. Under this situation, there is high possibility that the terrestrial DMB market will collapse. Accordingly, a policy helping the two services to form a complementary relationship with each other will be welcome for long-term development of this market.

The Relationship With Other Competitive Services

In Korea's communications service market, besides standard mobile wireless data services, there are various services. Among them, WiBro service recently received much attention from the public. It was launched in February 2005. WiBro service is an official name of Korea's portable Internet service. It is a new-concept, wireless high-speed Internet service using the 2.3GHz band, enabling indoor and outdoor reception by

devices moving even at nomadic speed. Many Korean telecommunication companies, including SK Telecom, KT, and Dacom, are applying for its license. Many experts forecast that its ripple effects are almost comparable to those caused by voice mobile communications services half a decade ago. However, a simultaneous launch of two similar services, DMB service and WiBro service, can reduce the possibility of successful market penetration in the short run. In spite of the risk, these new services will play a crucial role for the industry. By identifying the relationship between DMB service and WiBro service, we can investigate ways that make the two services more competitive. This session examines their relationships from three different areas of interest.

Service Coverage and Content Variety

In the case of WiBro, it is practically impossible to offer a nationwide coverage during the initial stage after the service launch. Although it has the national coverage at later stages, it still has shadow zones in urban downtown sections. In regards to DMB, satellite DMB can easily offer national coverage, whereas terrestrial DMB is limited to Seoul and some metropolitan areas. By using special transporters, both terrestrial DMB and WiBro can cover shadow zones (including downtown districts and underground) even during early stages of commercialization.

In terms of the variety of content provided, while WiBro service offers a full spectrum of contents, ranging from information to entertainment, DMB service focuses exclusively on multimedia entertainment programs, with only a limited segment dedicated to information such as weather forecast service in collaboration with the Korea Meteorological Administration. In terms of live broadcasting, WiBro, unable to provide real-time TV broadcasting, appears less competitive than DMB, transmitting live broadcasting and contents. So far, terrestrial TV content is only supplied through terrestrial DMB. Also, satellite DMB service is going to retransmit terrestrial TV contents. If both terrestrial DMB and satellite DMB retransmit regular terrestrial TV contents, WiBro service will be inferior to DMB service. According to a survey on DMB demand conducted by ETRI in 2004, 41.1% of sampled respondents said they would be interested in subscribing to a DMB service, if DMB retransmitted TV contents in addition to DMB's own programs. Only 17.1% of the respondents were interested in DMB, in the case where DMB provided only its own programming.

Figure 6. Preferences of WiBro service contents

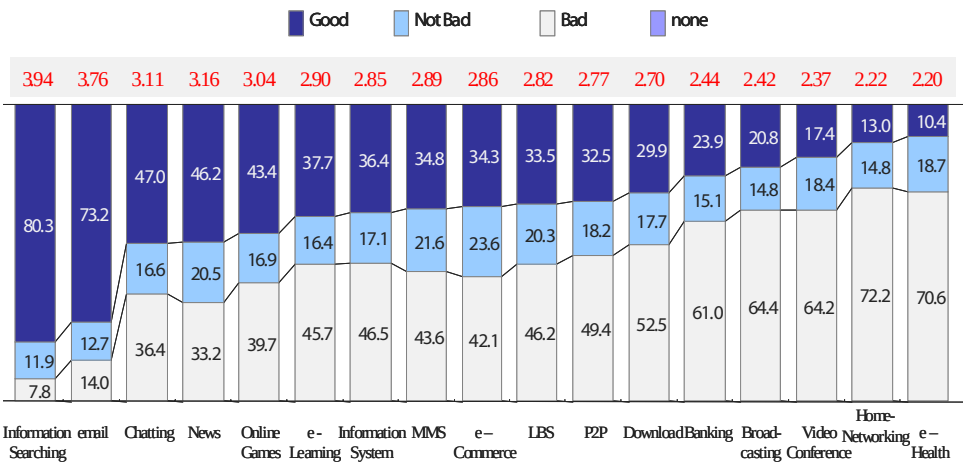
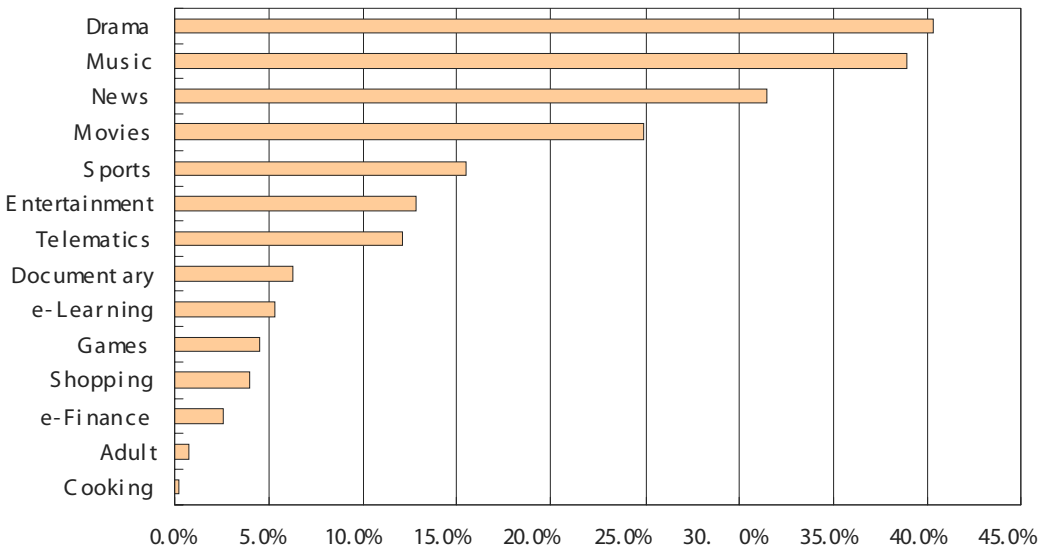


Figure 7. Preferences of DMB service contents



Customers

As can be seen in Figure 6, in the case of WiBro, the respondents of the ETRI survey (2004) expressed a higher degree of interest for information searching, e-mail, and news. By age group, the younger the respondents, the stronger the appeal for contents such as chatting, games, and messaging. In comparison, older age groups showed more interest in value-added services which could yield practical benefits.

Contrasting responses were obtained regarding DMB service (see Figure 7). Drama series, music, and news were the preferred types of content that attracted respondents toward DMB. E-finance and shopping-related contents appeared to weigh in relatively little in their interest in DMB. By age group, teenagers exhibited a strong preference for music and entertainment, while respondents in their 20s and early 30s mostly favored films and animations, and those in their mid-30s or older, were interested in sports broadcast.

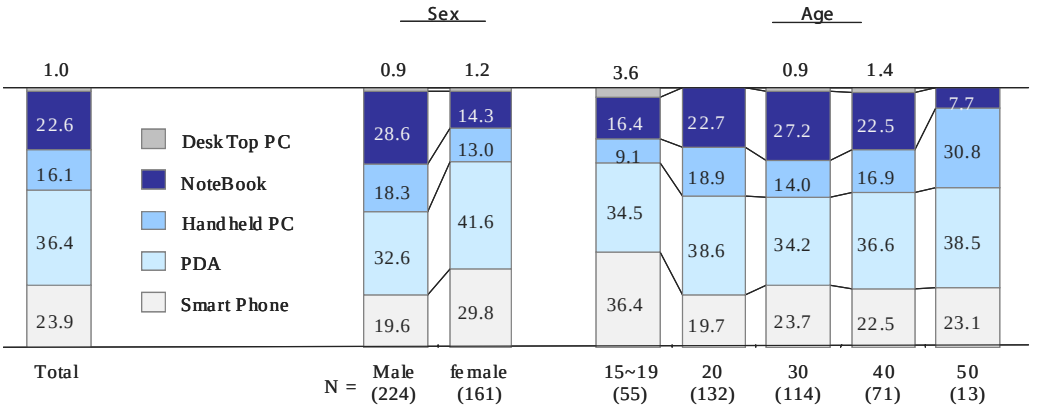
Receiving Devices

Figure 8 indicates that PDA and notebook computers were the preferred receiving devices for WiBro service across all age groups.

Regarding DMB, handheld devices (61.2%) and car-mounted devices (23.8%) were the preferred devices. Respondents showed only a low level of interest in stand-alone devices (4.4%). Meanwhile, an overwhelming majority of respondents favored a dual-reception device enabling the reception of both terrestrial and satellite DMB (80%). Moreover, as many as 74% answered they would wait with their DMB subscription until the dual-reception devices became available.

This survey finds several common features shared by the two services. First of all, they are both multimedia services, and both services target teens as their principal customer group. Moreover, they are portable services focusing on the customer group primarily interested in using the services under an outdoor and mobile environment. On the other hand, the survey results also indicate a number of differences. First,

Figure 8. WiBro service receiving device preference



WiBro service is perceived as an information and e-commerce-oriented service, in contrast to DMB, which is extensively associated with multimedia services. The preferred receiving devices for WiBro and DMB services were also clearly distinct. The second major difference has to do with service coverage. While WiBro service coverage is limited during the early stages of its implementation, satellite DMB can offer nation-wide coverage from the outset. Table 3 compares WiBro with DMB.

If terrestrial and satellite DMB launch as competing services, it is likely to reduce their appeal to customers and to ultimately weaken their competitiveness vis-à-vis other services. Furthermore, although two services partially overlap with each other in service content and target customer groups, it is not a major problem in building a collaborative relationship between the two services. This phenomenon has occurred in many other products and services. In fact, there is a high possibility that entirely new services or products turn out to be ones that fail to reflect market needs. In other words, when the DMB

Table 3. WiBro vs. DMB

		Category	
		WiBro	DMB
Service Coverage/ Content Variety	Coverage	Initial stage: 7 major cities Thereafter: Nationwide Preferred reception locations: Car, bus, train	Terrestrial DMB: Seoul and metropolitan areas Satellite DMB: Nationwide Preferred reception locations: Car, bus, train
	Content Variety	Information, e-commerce, chatting and communication, multimedia services	Multimedia-centered services
Customer		Prefer information and shopping services	Prefer multimedia services
Receiving Device		PDA, notebook computer	Mobile phone, car-mounted devices
Price		30,000 won	7,640 won

service is introduced to the market as a complementary service, rather than a competing service, to other similar services, its ripple effect over the communications market will be stronger and more substantial. Accordingly, for the successful launch of DMB and WiBro, instead of forming a rival relationship with WiBro service, DMB must be seen as complementing the perceived shortcoming of WiBro, namely its weakness in multimedia content offering. For long-term interest of the two services, it is wise for them to position themselves as each other's companion services.

POLICY ISSUES

Monopoly Market

Due to the withdrawal of bid by KT, the satellite DMB market has been monopolized by SK Telecom-affiliated TU Media. However, in the case of terrestrial DMB, many experts are still debating whether monopoly market or free competition market is appropriate. To prevent large broadcasting companies from monopolizing the terrestrial DMB market, the Broadcasting Act proposed by the Korea Broadcasting Commission limits the number of multiplex to one per operator. By using one multiplex, broadcasting companies can operate one video, three audio, and one data channels. If they want to operate only one multiplex, they can operate up to two video channels. However, KBS (Korean Broadcasting System, www.kbs.co.kr), the nation's representative public broadcaster, plans to operate two multiplexes for retransmission of KBS 1 TV, KBS 2 TV, and EBS (Education Broadcasting System). Other private broadcasting companies, including MBC (www.mbc.co.kr) and SBS (www.sbs.co.kr), are strongly protesting KBS's plan, by saying that if KBS operates two multiplexes and others operate one multiplex, it will constitute an unfair preferential treatment. Industry insiders are expressing apprehension as to the possibility of this discord leading to an outbreak of large controversy on an alleged inequity. The Korean Government is still investigating appropriate market structures of terrestrial DMB.

Retransmission of Terrestrial TV

In June 2004, at a public hearing, the Korea Broadcasting Commission disclosed a regulatory decision prohibiting the retransmission of terrestrial TV programs by satellite DMB, with the exception of KBS 1 TV and EBS. TU Media, a satellite DMB operator, was fiercely protesting against the prohibition of terrestrial TV retransmission. Numerous consumer surveys indicate that potential viewers are substantially less interested in subscribing to satellite DMB, if this service does not transmit terrestrial TV. Accordingly, with this decision by the Government, the market prospect of satellite DMB business as a whole will be put into question. Recently, Skylife, a Korean satellite broadcaster, finally received the authorization to retransmit terrestrial TV (MBC, SBS) starting next year. This restriction has been a true obstacle for satellite DMB providers' efforts to expand subscriber pools, and continues to limit and mitigate the outlook for this business area. Because the Korean Government is trying to prevent satellite DMB market from becoming a monopoly, its decision will not change anytime soon.

On the other hand, if the retransmission of terrestrial TV by satellite DMB is permitted, it will cancel out the key competitive elements of terrestrial DMB. Accordingly, it is advisable that the Government, rather than making an unconditional lifting of the retransmission restriction, instead wait and see how the market plays out and adjust its policy accordingly, so that these two DMB services can develop a complementary relationship.

Channel Compositions

The 25MHz frequency band used by satellite DMB has a transmission capacity of 7Mbps, and it is sufficient for operating up to 13 video channels. As public interests in satellite DMB are concentrated on video contents rather than audio contents, video content transmission is a priority for satellite DMB providers. However, because the Korean government strongly believes that terrestrial DMB service is inferior to satellite DMB, it supports terrestrial DMB and imposes many restrictions on satellite DMB. Given that situation, if the relationship between satellite DMB and terrestrial DMB turns out to be competitive rather than complementary, the restrictions on satellite DMB in the number of video channels allowed would be unfair and inappropriate. The overall DMB service market will be better served by loosening regulatory, directly geared toward stimulating terrestrial DMB business, rather than burdening the competing services with further restrictions such as a video channel restriction for satellite DMB.

CONCLUSIONS

This chapter reviews DMB technology, its business models, its market structures, and its policies. In particular, it explores business opportunities around DMB services and identifies major issues that must be solved to launch DMB services successfully. While the convergence of broadcasting and telecommunications is still in its early stages, it is nevertheless useful to draw a few preliminary conclusions from this survey in order to forecast trends in the worldwide telecommunication market, as well as the Korean telecommunication market.

In 1995–1997, many European countries, such as the UK, Sweden, France, and Germany, introduced DAB being similar to DMB into their markets. However, their markets have not expanded quickly. Experts point out that this market failure is caused by inability of existing terrestrial broadcasting companies in marketing and new service developments, and their conservative investments. Because telecommunication companies that have ample funds and are experienced in aggressive investment and marketing, play important roles in the DMB business in Korea, they are expected to overcome the problems encountered in European countries. The future of DMB in Korea is unpredictable, as it introduces a range of interrelated political, economic, and technical challenges. Particularly, the repeated failures in establishing a regulatory framework for DMB and ill-defined roles of players in the DMB industry might interfere the diffusion of DMB in the Korean market. As the convergence of broadcasting and telecommunications makes progress, proper modifications of existing regulatory frameworks should be made in order to guarantee success of DMB service in Korea.

REFERENCES

ETRI. (2004). *Survey for DMB service* (Working paper).

Simpson, S. (2004). Universal service issues in converging communications environments: The case of the UK. *Telecommunication Policy, 28*, 233–248.

SK Telecom. (2003). Satellite DMB business plan (Working paper).

Tadayoni, R., & Skouby, K.E. (1999). Terrestrial digital broadcasting: Convergence and its regulatory implication. *Telecommunication Policy, 23*, 175–199.

Wu, I. (2004). Canada, South Korea, Netherlands and Sweden: Regulatory implications of the convergence of telecommunications, broadcasting and Internet services. *Telecommunications Policy, 28*, 79–96.

Yu, B. (2003). Policy directions for introduction of terrestrial digital multimedia broadcasting. KT Telecommunication Market.

This work was previously published in Unwired Business: Cases in Mobile Business, edited by S. J. Barnes and E. Scornavacca, pp. 270-284, copyright 2006 by IRM Press (an imprint of IGI Global).

Section 7

Critical Issues

This section addresses conceptual and theoretical issues related to the field of multimedia technologies, which include quality of service issues in multimedia transmission and the numerous approaches adopted by researchers that aid in making multimedia technologies more effective. Within these chapters, the reader is presented with analysis of the most current and relevant conceptual inquires within this growing field of study. Particular chapters address methodologies for the organization of multimedia objects and the relationship between cognition and multimedia. Overall, contributions within this section ask unique, often theoretical questions related to the study of multimedia technologies and, more often than not, conclude that solutions are both numerous and contradictory.

Chapter 7.1
Perceived Quality Evaluation for Multimedia Services

H. Koumaras
University of Athens, Greece

E. Pallis
Technological Educational Institute of Crete, Greece

G. Xilouris
University of Athens, Greece

A. Kourtis
N.C.S.R., Demokritos, Greece

D. Martakos
University of Athens, Greece

INTRODUCTION

The advent of 3G mobile communication networks has caused the fading of the classical boundaries between telecommunications, multimedia, and information technology sectors. The outcome of this convergence is the creation of a single platform that will allow ubiquitous access to the Internet, multimedia services, and interactive audiovisual services, and in addition (and most importantly) offering the required/appropriate perceived quality level at the end user's premises.

In this respect, multimedia services that distribute audiovisual content over 3G/4G mobile communication systems are expected to posses a major part of the bandwidth consumption, making necessary the use of video compression. Therefore, encoding techniques (e.g., MPEG, H-26x) will be applied which achieve high compression ratios by exploiting the redundancy in the spatiotemporal domain of the video content, but as a consequence produce image artifacts, which result in perceived quality degradation.

One of the 3G/4G visions is the provision of audiovisual content at various quality and price levels. There are many approaches to this issue, one being the perceived quality of service (PQoS) concept. The evaluation of the PQoS for audiovisual content will provide a user with a range of potential choices, covering the possibilities of

low-, medium-, or high-quality levels. Moreover the PQoS evaluation gives the service provider and network operator the capability to minimize the storage and network resources by allocating only the resources that are sufficient to maintain a specific level of user satisfaction.

The evaluation of the PQoS is a matter of post-encoding procedures. The methods and techniques that have been proposed in the bibliography mainly aim at:

- determining the encoding settings (i.e., resolution, frame rate, bit rate) that are required in order to carry out successfully a communication task of a multimedia application (i.e., videoconference); and
- evaluating the quality level of a media clip based on the detection of artifacts on the signal caused by the encoding process.

The scope of this article is to outline the existing procedures and methods for estimating the PQoS level of a multimedia service.

BACKGROUND

The advent of quality evaluation was based on applying pure mathematical/error-sensitive equations between the encoding and the original/uncompressed video signal. These primitive methods, although they provided a quantitative approach about the quality degradation of the encoded signal, do not provide reliable measurements of the perceived quality, because they miss the characteristics and sensitivities of the human visual system.

The most widely used primitive methods and quality metrics that are based on the error sensitivity framework are the peak signal to noise ratio (PSNR) and the mean square error (MSE):

$$\text{PSNR}=10\log_{10}\frac{L^2}{\text{MSE}}$$, *where L denotes the dynamic pixel value (i.e., equal to 255 for 8bits/pixel monotonic signal)* (1)

$$\text{MSE}=\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2$$, *where N denotes the total pixels and* x_i/y_i *the* i^{th} *pixel value in the original/distorted signal* (2)

Currently, the evaluation of the PQoS is a matter of objective and subjective evaluation procedures, each time taking place after the encoding process (post-encoding evaluation). Subjective picture/audio quality evaluation methods require a large amount of human resources, establishing it as a time-consuming process (e.g., large audiences evaluating video/audio sequences). Objective evaluation methods, on the other hand, can provide PQoS evaluation results faster, but require a large amount of machine resources and sophisticated apparatus configurations. Towards this, objective evaluation methods are based on and make use of multiple metrics, which are related to the content's artifacts (i.e., tilling, blurriness, error blocks, etc.) resulting during an encoding process.

These two categories of PQoS evaluation methods will be analyzed and discussed in the following sections.

SUBJECTIVE QUALITY EVALUATION METHODS

The subjective test methods, which have mainly been proposed by the International Telecommunications Union (ITU) and the Video Quality Experts Group (VQEG), involve an audience of people who watch a video sequence and score its quality, as perceived by them, under specific and controlled watching conditions. Afterwards, the statistical analysis of the collected data is used

for the evaluation of the perceived quality. The mean opinion score (MOS) is regarded as the most reliable subjective metric of quality measurement and has been applied on the most known subjective techniques.

Subjective test methods are described in ITU-R Rec. T.500-11 (2002) and ITU-T Rec. P.910 (1999), suggesting specific viewing conditions, criteria for the observer, test material selection, assessment procedure description, and statistical analysis methods. The BT.500-11 describes subjective methods that are specialized for television applications, whereas ITU-T Rec. P.910 is intended for multimedia applications.

The most known and most widely used subjective methods are:

- **Double Stimulus Impairment Scale (DSIS):** This method proposes that observers watch multiple references and degraded scene pairs, with the reference scene always shown first. Scoring is evaluated on an overall impression scale of impairment: imperceptible, perceptible but not annoying, slightly annoying, annoying, and very annoying. This scale is commonly known as the five-point scale (where 5 corresponds to "imperceptible" and 1 to "very annoying").
- **Single Stimulus (SS) Methods:** Multiple separate scenes are shown. There are two different SS approaches: SS with single view of test scenes and SS where the test scenes are repeated. Three different scoring methods are used:
  - **Adjectival:** The aforementioned five-grade impairment scale, however half-grades are allowed.
  - **Numerical:** An 11-grade numerical scale, useful if a reference is not available.
  - **Non-Categorical:** A continuous scale with no numbers or a large range, for example, 0-100.
- **Stimulus Comparison Method:** This methods exploits two well-matched screens, where the differences between scene pairs are scored in one of the two following scoring methods:
  - **Adjectival:** A seven-grade, +3 to -3 scale labeled: much better, better, slightly better, the same, slightly worse, worse, and much worse.
  - **Non-Categorical:** A continuous scale with no numbers or a relation number either in absolute terms or related to a standard pair.
- **Single Stimulus Continuous Quality Evaluation (SSCQE):** According to this method, the viewers watch a program of typically 20-30 minutes without any reference signal. The viewers, using a slider, continuously rate the instantaneously perceived quality using an adjectival scale from 'bad' to 'excellent', which corresponds to an equivalent numerical scale from 0 to 100.
- **Double Stimulus Continuous Quality Scale (DSCQS):** At DSCQS the viewers watch multiple pairs of quite short (i.e., 10 seconds) reference and test sequences. Each pair appears twice, with random order of the reference and the test sequence. The viewers/subjects are not aware of the reference/test order, and they are asked to rate each of the two separately on a continuous adjectival scale, ranging from 'bad' to 'excellent', which corresponds to an equivalent numerical scale from 0 to 100. This method is usually used for evaluating slight quality differences between the test and the reference sequence.

The aforementioned methods are described in the ITU-R Rec. T.500-11 document and are mainly intended for television signals. Based on slight modifications and adaptations of these methods, some other subjective evaluation methods (namely absolute category rating (ACR), degradation cat-

egory rating (DCR), etc.) for multimedia services are described in ITU-T Rec. P.910.

OBJECTIVE QUALITY EVALUATION METHODS

The preparation and execution of subjective tests is costly and time consuming, and its implementation today is limited to scientific purposes, especially at VQEG experiments.

For this reason, a lot of effort has recently been focused on developing cheaper, faster, and more easily applicable objective evaluation methods. These techniques successfully emulate the subjective quality assessment results, based on criteria and metrics that can be measured objectively. The objective methods are classified according to the availability of the original video signal, which is considered to be of high quality.

The majority of the proposed objective methods in the literature require the undistorted source video sequence as a reference entity in the quality evaluation process, and due to this are characterized as full reference methods. These methods emulate characteristics of the human visual system (HVS) using contrast sensitivity functions (CSF), channel decomposition, error normalization, weighting, and finally Minkowski error pooling for combining the error measurements into single perceived quality estimation (Wang, Sheikh, & Bovik, 2003).

However it has been reported (VQEG, 2000; Wang, Bovik, & Lu 2002) that these complicated methods do not provide more accurate results than the simple mathematical measures (such as PSNR). Due to this some new full reference metrics that are based on the video structural distortion, and not on error measurement, have been proposed (Wang et al., 2003).

On the other hand, the fact that these methods require the original video signal as reference deprives their use in commercial video service applications, where the initial undistorted clips are not accessible. Moreover, even if the reference clip is available, then synchronization predicaments between the undistorted and the distorted signal (which may have experienced frame loss) make the implementation of the full reference methods difficult and impractical.

Due to these reasons, recent research has focused on developing methods that can evaluate the PQoS level based on metrics, which use only some extracted structural features from the original signal (Reduced Reference Methods—Guawan & Ghanbari, 2003) or do not require any reference video signal (No Reference Methods—Lu, Wang, Bovik, & Kouloheris, 2002).

However, due to the fact that the 3G/4G vision is the provision of audiovisual content at various quality and price levels (Seeling, Reisslein, & Kulapala, 2004), there is great need for developing methods and tools that will help service providers to predict quickly and easily the PQoS level of a media clip. These methods will enable the determination of the specific encoding parameters that will satisfy a certain quality level. All the previously mentioned post-encoding methods may require repeating tests in order to determine the encoding parameters that satisfy a specific level of user satisfaction. This procedure is time consuming, complex, and impractical for implementation on the 3G/4G multimedia mobile applications.

Towards this, recently research was performed in the field of pre-encoding estimation and prediction of the PQoS level of a multimedia service as a function of the selected resolution and the encoding bit rate (Koumaras, Kourtis, & Martakos, 2005; Koumaras et al., 2004). These methods provide fast and quantified estimation of the PQoS, taking into account the instant PQoS variation due to the spatial and temporal (S-T) activity within a given encoded sequence. Quantifying this variation by the mean PQoS (MPQoS) as a function of the video encoding rate and the picture resolution, it finally used the MPQoS as a metric for pre-encoding PQoS assessment based

on the fast estimation of the S-T activity level of a video signal.

FUTURE TRENDS

Simultaneously with the development of the aforementioned methods and techniques, research has been focused on developing methods that determine the adequate quality level for a specific multimedia application, taking under consideration not solely the visual estimations, but also a great number of parameters and metrics that depend on the task nature and the user emotional behavior and psychophysical characteristics (Mullin, Smallwood, Watson, & Wilson, 2001). For example, the classification of the task as foreground or background in correlation with its complexity (Buxton, 1995) is a parameter that differentiates the quality demands of a multimedia application. On the other hand, the emotional content of a multimedia communication task alters the required quality level of the specific communication service (Olson, 1994). Due to this, various parameters are measured in order to estimate the appropriate minimum quality level of a multimedia application. Such parameters are:

- the user characteristics (i.e., knowledge background, language background, familiarity with the task, age);
- the situation characteristics (i.e., geographical remoteness, simultaneous number of users, distribution of users);
- the user cost (i.e., heart rate, blood volume pulse); and
- the user behavior (i.e., eye tracking, head movement).

However, these methods still have some issues to solve on the technical, theoretical, and practical levels. A user that participates in such an assessment procedure is so wired (even on the head, he or she may wear the eye tracking equipment) that it causes uncomfortable feelings and affects his or natural behavior. Technical issues, such as the eye tracking loss and the manual calibration/correction by a human operator, affect the reliability of the methods in real-time environments (Mullin et al., 2001).

CONCLUSION

Multimedia applications, and especially encoded video services, are expected to play a major role in third-generation (3G) and beyond mobile communication systems. Given that future service providers are expected to provide video applications at various price and quality levels, quick and economically affordable methods for preparing/encoding the offering media at various qualities need to be developed. There are a number of approaches to this challenge, one being the use of the perceived quality of service concept. The evaluation of the PQoS for multimedia and audiovisual content that has variable bandwidth demands will provide a user with a range of choices covering the possibilities of low-, medium-, or high-quality connections, an indication of service availability and cost. This article outlines the various existing PQoS evaluation methods and comments on their efficiency.

These methods can be mainly categorized into two major classes: subjective and objective. The subjective test methods involve an audience of people who watch a video sequence and evaluate its quality, as perceived by them, under specific and controlled watching conditions. The objective methods successfully emulate the subjective quality assessment results, based on criteria and metrics that can be measured objectively. These objective methods are classified according to the availability of the original video signal to full reference, reduced reference, and no reference.

However, all the aforementioned post-encoding methods require repeating post-encoding tests in order to determine the encoding parameters that

satisfy a specific level of user satisfaction, making them time consuming, complex, and impractical for implementation on the 3G/4G multimedia mobile applications. Due to this, lately some new pre-encoding evaluation methods have been proposed that are capable of estimating/predicting the PQoS level of a multimedia service based on the selected resolution, bit rate, and content activity. These methods quickly provide accurate estimations of PQoS level, alleviating the time and resource requirements that the traditional objective methods consume.

ACKNOWLEDGMENT

The work in this article was carried out in the frame of the Information Society Technologies (IST) project ENTHRONE/FP6-507637.

REFERENCES

Buxton, W. (1995) Integrating the periphery and context: A new taxonomy of telematics. *Proceedings of Graphics Interface 1995* (pp. 239-246).

Guawan, I. P., & Ghanbari, M. (2003). Reduced-reference picture quality estimation by using local harmonic amplitude information. *Proceedings of the London Communications Symposium 2003.*

Koumaras, H., Kourtis, A., & Martakos, D. (2005). Evaluation of video quality based on objectively estimated metric. *Journal of Communications and Networking, 7*(3), 235-242.

Koumaras, H., Pallis, E., Xilouris, G., Kourtis, A., Martakos, D., & Lauterjung, J. (2004). Pre-encoding PQoS assessment method for optimized resource utilization. *Proceedings of the 2nd International Conference on Performance Modeling and Evaluation of Heterogeneous Networks (Het-NeTs04),* Ilkley, UK.

Lu, L., Wang, Z., Bovik, A. C., & Kouloheris, J. (2002). Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video. *Proceedings of the IEEE International Conference on Multimedia.*

Mullin, J., Smallwood, L., Watson, A., & Wilson, G. (2001). New techniques for assessing audio and video quality in real-time interactive communications. *Proceedings of the 3rd International Workshop on Human Computer Interaction with Mobile Devices,* Lille, France.

Olson, J. (1994). In a framework about task-technology fit, what are the tasks features? *Proceedings of CSCW '94: Workshop on Video Mediated Communication: Testing, Evaluation & Design Implications.*

Seeling, P., Reisslein, M., & Kulapala, B. (2004). Network performance evaluation using frame size and quality traces of single layer and two layer video: A tutorial. *IEEE Communications Surveys, 6*(3).

VQEG. (2000). *Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment.* Retrieved from http://www.vqeg.org

Wang, Z., Bovik, A. C., & Lu, L. (2002). Why is image quality assessment so difficult? *Proceedings of the IEEE International Conference in Acoustics, Speech and Signal Processing* (Vol. 4, pp. 3313-3316).

Wang, Z., Sheikh, H. R., & Bovik, A. C. (2003). Objective video quality assessment. In B. Furht & O. Marqure (Eds.), *The handbook of video databases: Design and applications* (pp. 1041-1078). CRC Press.

KEY TERMS

Bit Rate: A data rate expressed in bits per second. In video encoding the bit rate can be *constant*, which means that it retains a specific value for the whole encoding process, or *variable*, which means that it fluctuates around a specific value according to the content of the video signal.

Double Stimulus Continuous Quality Scale (DSCQS): A subjective evaluation method according to which the viewers watch multiple pairs of quite short (i.e., 10 seconds) reference and test sequences. Each pair appears twice, with random order of the reference and the test sequence.

Multimedia: The several different media types (e.g., text, audio, graphics, animation, video).

Objective Measurement of Perceived Quality: A category of assessment methods that evaluates the PQoS level based on metrics, which can be measured objectively.

Perceived Quality of Service (PQoS): The perceived quality level that a user experiences from a multimedia service.

Quality Degradation: The drop of the perceived quality to a lower level.

Single Stimulus Continuous Quality Evaluation (SSCQE): A subjective evaluation method according to which the viewers watch a program of typically 20-30 minutes, without the original reference shown, and score its quality.

Spatial-Temporal Activity Level: The dynamics of the video content, in respect to its spatial and temporal characteristics.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 758-762, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.2
Distributed Approach for QoS Guarantee to Wireless Multimedia

Kumar S. Chetan
NetDevices India Pvt Ltd, India

P. Venkataram
Indian Institute of Science, India

Ranapratap Sircar
Wipro Technologies, India

INTRODUCTION

Providing support for QoS at the MAC layer in the IEEE 802.11 is one of the very active research areas. There are various methods that are being worked out to achieve QoS at MAC level. In this article we describe a proposed enhancement to the DCF (distributed coordination function) access method to provide QoS guarantee for wireless multimedia applications.

Wireless Multimedia Applications

With the advancement in wireless communication networks and portable computing technologies, the transport of real-time multimedia traffic over the wireless channel provides new services to the users. Transport is challenging due to the severe resource constraints of the wireless link and mobility. Key characteristics of multimedia-type application service are that they require different quality of service (QoS) guarantees.

The following characteristics of WLAN add to the design challenge:

- Low bandwidth of a few Mbps compared to wired LANs bandwidth of tens or hundreds Mbps.
- Communication range is limited to a few hundred feet.
- Noisy environment that leads to high probability of message loss.
- Co-existence with other potential WLANs competing on the same communication channel.

Figure 1. Block schematic of proposed system at AP

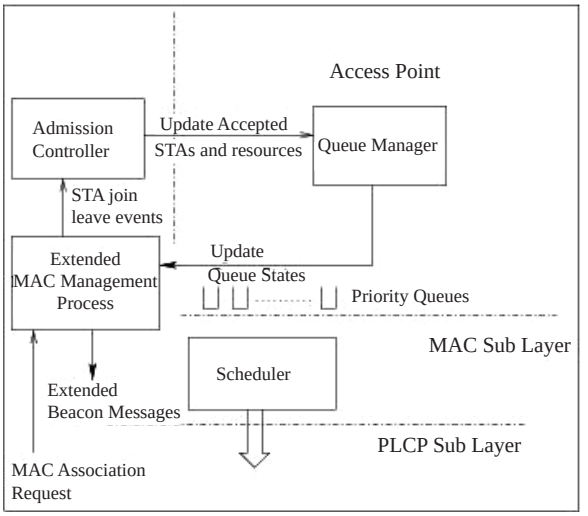
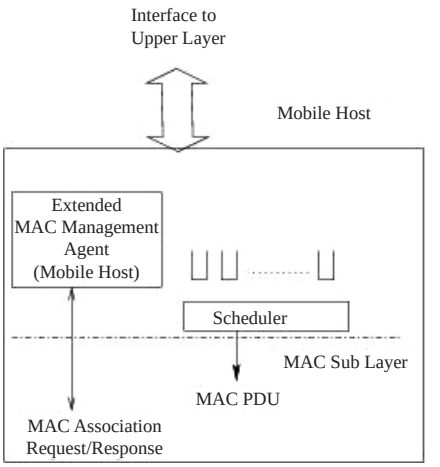


Figure 2. Block schematic of proposed system at mobile station



Successful launching of multimedia applications requires satisfying the application's QoS requirements.

The main metrics (or constraints) mentioned in such guidelines and that eventually influence the MAC design are: time delay, time delay variation and data rate. We develop a scheme to provide guaranteed data rate for different applications in WLAN environment.

PROPOSED ENHANCEMENT OF DCF TO PROVIDE QoS

The proposed enhancement is developed as a modular system, which integrates with DCF MAC of 802.11b wireless LAN.

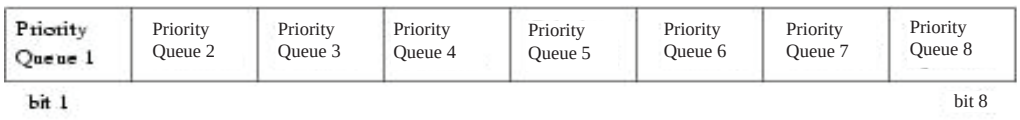
Salient Features of the Modular System

- Provides throughput guarantee for traffic flow between a pair of mobile stations.
- Works in distributed mode.
- Provides MAC level admission control for traffic flow.
- Applications on the mobile stations can send resource reservations request for each call (session).
- Works with backward compatibility, on hosts that do not support QoS enhancements.

Based on the basic principle of DCF access mode, each mobile station transmits data independent of other mobile station. Also, the AP (access point) has no role to play during the data transmission. Under this scenario, the throughput control (and guarantee) has to be achieved in a distributed manner.

One has to restrain a station from accessing the medium if there are other stations in the BSS that has requested for higher resource. If there is no such other station, the station is allowed to access the medium. We propose to use eight different priority flows. The queue manager at mobile stations maintains queues for these flows. Also, the state of these queues (if there are applications that are using this flow) is synchronized across all

Figure 3. Queue state field



Bit = 1 Queue is active
Bit = 0 Queue is inactive

Algorithm 1. Call admission controller algorithm

```
Begin
        for ( i = Min_Priority;i <= Max_Priority; i++ )
                do
                        if ( Requested_Tput + Current_Tput[i] <= Max_Tput[i] )
                                then
                                        goto accept_call
                                else
                                        continue
                        endif
                done
        Call RejectCall() /* No priority flow could fit this call */
        goto end

        accept_call:
                Current_Tput[i] = Current_Tput[i] + Request_Tput
                Call AcceptCall()
End
```

the mobile station via the beacon messages sent by AP. The scheduler transmits the packets from these priority queues using a priority algorithm. The admission controller admits a call to particular flow, if the acceptance of call to the flow do not over-shoot the throughput for that flow.

ARCHITECTURE OF THE MODULAR SYSTEM

The block schematics of the proposed system are given in Figures 1 and 2, for AP and mobile station respectively. The system has four major components:

1. Extended MAC management process
2. Admission controller
3. Queue manager
4. Scheduler

The detailed functioning of each of the components is explained in the following subsections.

Enhanced MAC Management Process

To signal the QoS messages, we propose an extension to MAC layer management messages to carry the resource request and responses.

To signal mobile host resource requirements to AP, we propose extension to the existing MAC management frames. This approach does

Algorithm 2. Scheduling packets

```
begin
        r  = rand()    /* r[0,1] */
        CurrentProb   0
        for(i=1 to N) /* N is number of active flows */
                do
                        if (r < CurrentProb)
                                Call TransmitPacket(i)
                        else
                                CurrentProb  = CurrentProb+NormPriority[i]
                        endif
                done
end
```

not need any changes in the core MAC layer. The SME (station management entity), which is normally residing in a separate management plane, needs modifications, which can be easily incorporated.

We incorporate enhanced MAC management with two segments.

MAC Management Process at AP

Apart from receiving and transmitting extended management frames to signal QoS, here the AP also broadcasts the queue states as an extension to beacon message.

MAC Management agent at Mobile Host

Apart from normal functionalities, the agent also receives the extended beacon message with queue states and passes this information to QM.

Management Message Modified

The beacon message is extended to include the queue states in a bit mask format. The queue state is an eight-bit field with each field representing a priority queue state. The queue is considered active if the bit is set, else inactive.

Admission Controller

The admission controller gets triggered by the extended MAC management process. The admission controller is a parameter based admission controller. However the decision process is modified to suit the distributed nature of DCF functionalities. While working in DCF mode each mobile station transmits/receives PDUs independent of AP and independent of other mobile stations. However the queue manager and scheduler as designed such that each station (while being completely independent) will transmit the PDUs at certain priority class. In order for the admission controller to admit a call, it needs to identify the maximum available throughput for a particular flow.

Based on [3] and [4] maximum achievable throughput per priority class is estimated as follows.

$$\text{Max Tput}[i] = \frac{P(\text{successful tx} \mid \text{flow} = i) * \text{Data pay load size}}{P(\text{collision}) * \text{Dur}_{\text{collision}} + P(\text{slot is idle}) * \text{aSlotTime} + P(\text{successful tx}) * \text{Dur}_{\text{success}}} \quad (1)$$

where $\text{Dur}_{\text{success}}$ is time duration for successful transmission of PDU, $\text{Dur}_{\text{collision}}$ is collision duration and aSlotTime is duration of one slot interval.

When a new call request arrives, the call is tried to fit in a flow of highest priority. By doing so, if the call requested throughput is achieved and existing calls are not disturbed, the call is accepted, or else the priority is decreased till the call is acceptable, if the call is not acceptable beyond least priority, the call is rejected. The admission control algorithm is described as follows.

Queue Manager

The main functionality of the queue manager is to synchronize the queue states across all mobile stations. Due to lack of any centralized coordinator, the queue manager synchronizes the states using broadcast messages. The queue manager at the AP broadcasts the states of queues to all the mobile stations in the BSS using extension to Beacon frames.

A queue for a particular flow is active if there are any calls in that flow, otherwise the queue is inactive. The queue states are broadcasted using bit masks in the extended beacon message, with each bit representing a flow. If the bit is set to 1, the queue for that flow is active; otherwise the queue is inactive. The queue manager module on station receives these broadcast messages and updates the queue state.

When all the stations have exchanged the states of the queue, each of the mobile stations would have synchronized queue states. Thus all the stations in the BSS will have a queue at particular priority level as active.

Scheduler

The scheduler's job to pick a packet from the priority flow queue and schedule them for transmission. The scheduler picks the packet from active queues. Irrespective of the fact the queue has data or not, the PacketTransmit function is called by the scheduler. The scheduler picks up the packet randomly from any of the active queues only. Now the probability to choose a queue is equal to normalized value of the priority of the queue. We define the normalized priority P_n as

$$P_{ni} = \frac{P_i}{\sum_K P_k} \tag{2}$$

where P_i is the priority of i^{th} queue.

We have assumed equal sized packets in all queues. For non-equal packet sizes the normalized priority would become:

$$P_{ni} = \frac{\frac{P_i}{S_i}}{\sum_k \frac{P_k}{S_k}} \tag{3}$$

where S_k is the size of k_{th} packet.

Each of the active queues is arranged in the ascending order of the priority. The scheduler selects the queue to be scheduled, and send the packet for transmission. The scheduling algorithm is described in Algorithm 2.

Packet Transmission Process

The PacketTransmit function checks for the packet and hands over the packet to MAC layer if the queue has data.

If the queue is empty, no packet is passed to the MAC function; however the packet transmission will back-off for other stations to transmit. A timer call back for a duration that represents the time taken by an average-sized packet is registered. This will provide opportunity for other stations with the packet in the same priority level queue to transmit the packet on to the medium.

The MAC will perform DCF access method to access the medium and make transmission. If the packet could not be transmitted due to wireless medium loss or collusion, the packet is retransmitted by the MAC layer, the scheduling is not changed due to retransmission. The packet transmission algorithm is described in Algorithm 3.

Table 1. Simulation parameters

Data Rate	11 Mbps
RTS Threshold	3000
Physical Layer Frequency	2457e+6
Transmission Power	31 mW
Receiver Threshold	1.15209e-10
Radio Propagation Model	TwoRayGround
LongRetryLimit	2

FUNCTIONING OF THE MODULAR SYSTEM

The functions of modular system are to gather the QoS requirements from the mobile stations and schedule their applications as per the specified QoS.

The functioning of the system is explained by considering the following cases.

Case 1. When New Mobile Station Enters the BSS with New Application

When a new station enters the BSS, the station joins the BSS using normal association procedure (as per IEEE802.11 standard). The AP instantiates a MAC management agent. The agent is migrated to the mobile host. The station uses the "Extended MAC Management Agent" to signal the resource requirements for new application. The AP receives the resource request, and passes on the QoS parameters to admission controller. The admission controller examines the resource requests, applies the admission control algorithm and accepts/rejects the request. The accept/reject response is conveyed back to the mobile station via the "Extended MAC Management process" module and updates the queue states in the QM. The QM updates the queue states for each of the flows, and sends them to the mobile stations as an extension to beacon message.

Figure 4. Throughput plot for mobile station with CBR traffic under DCF mode

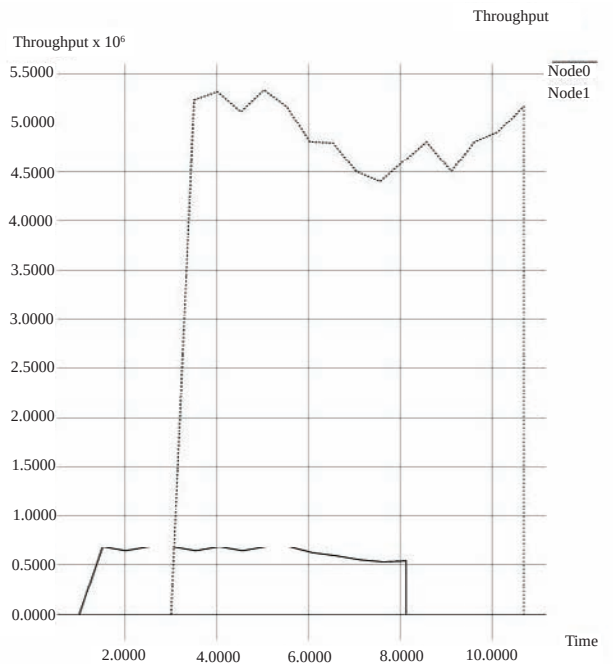


Figure 5. Throughput plot for mobile station with VBR traffic under DCF mode

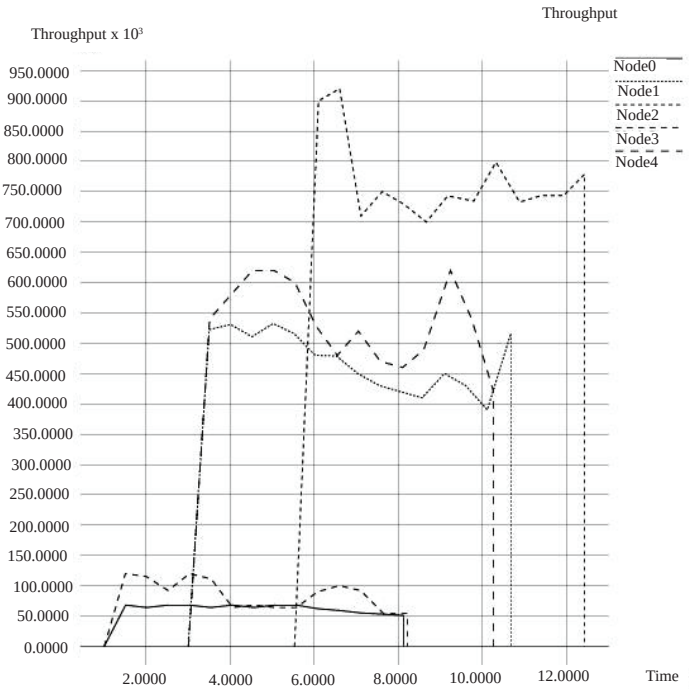
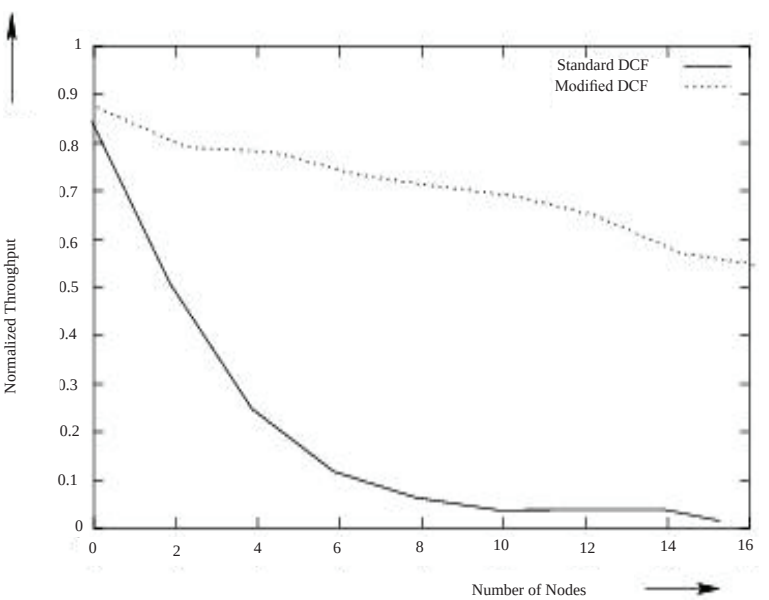


Figure 6. Normalized throughput for proposed scheme and standard DCF mod



Upon reception of the response for the call request, the mobile station can start sending traffic. The QM at the mobile station receives the queue states in the extended beacon message and updates local queue state. The scheduler the mobile station schedules the traffic using the queue state information.

Case 2. When a Mobile Station Leaves the BSS

When the mobile station is about leave the BSS, the station uses the enhanced MAC management process to signal the AP. The AP receives the station ID, and updates the QM and admission control to relinquish the resource used all application under this station and changes the queue state if required.

Case 3. When New Application Has to be Admitted

A mobile station is already associated with the AP, and may be running some applications or otherwise. Under this condition, the station and AP follow the normal steps (as in Case 1), as this is no different from Case 1.

Case 4. When Applications are Terminated at the Mobile Station

A mobile station stops the current application, but the station is still in the BSS. When the application is stopped, the station uses the enhanced MAC management process to signal the AP that the application has stopped. The AP follows steps in Case 2 to relinquish the resource and update queue state.

IMPLEMENTATION CONSIDERATION

The system has two components, that is to say, AP component and mobile station component.

- The AC is implemented at the centralized location, normally co-located with the AP.
- Extension to MAC management layer, QM and scheduler are implemented in both AP and mobile station.

AP Components Implementation

It is easier to enhance the AP to support functionalities such as CAC, QM and scheduler since there is relatively very few numbers of APs and also the APs are normally controlled by the service provider, who wishes to provide QoS in the BSS. The AC, QM and scheduler are implemented in high-level system language (C) and integrated into the AP software. The MLME functions on the AP are modified to understand the extensions in the MAC frames and pass the QoS parameters to the CAC. The receive (RX) and transmit (TX) functions on the AP are modified to call the QM and scheduler functions respectively.

Mobile Station Components Implementation

The station components are implemented as Java byte-code, and the agent is instantiated at the AP and migrated to the mobile station, as the mobile station gets associated with the AP. The extended MAC management agent, QM and scheduler are implemented in Java.

SIMULATION AND RESULTS

We have used ns-2 [7] as simulator to validate the proposed system. Ns are a discrete event simulator targeted at networking research. Ns provides substantial support for simulation of TCP, routing, and multicast protocols over wired and wireless (local and satellite) networks.

Simulation Setup and Procedure

Using ns-2, a sixteen-node network working in BSS and iBSS modes is setup. We have modified the MAC layer functionalities in ns-2 to support the proposed system.

In the simulated system, the mobile nodes communicate with each other and external world over the 802.11b 11MBPS channel. The transmission power of the mobile is chosen such that stations are not hidden from one another. If not stated otherwise, during the simulation neither RTS/CTS nor the fragmentation is used. For most of our simulation we have considered on-off traffic with exponential distribution, and log-normal distribution for the radio channel errors. We have done several experiments by considering traffic in both upstream and downstream direction. The allowed traffic has been classified as VBR (variable bit-rate) and CBR (constant bit-rate) with random arrivals. The duration of each experiment varies from 100 seconds to 1,000 seconds. The simulation parameters that are significant are listed in Table 1.

We have simulated the proposed systems as three individual units. We explain the simulation for each case with CBR and VBR traffics.

The DCF MAC functions at each station are modified according to the proposed scheme. Each of the mobile station generates traffic destined to other nodes with data packets of 1,500bytes, 1,000 bytes, 512 bytes and 64 bytes.

In our first experiment, we considered two nodes generating CBR traffic. Each node requested a throughput of 10\% and 90\% of the net throughput. In Figure 4, we have plotted the throughput vs. time for two nodes. It can be observed that both the stations get the allocated bandwidth. Also when the node0 application is terminated, the node1 uses up the excess bandwidth.

In the next experiment, we have considered BSS with five nodes admitted for scheduling. The nodes requested throughput requirements are at 50, 100, 500, 550, and 750 Kbps. In Figure 5, we have plotted the throughput versus time for all the five nodes. It can be observed from the plot, each of the nodes is scheduled to get the allocated bandwidth.

Comparison with the Standard DCF Scheme

Again we use the normalized throughput as a measure to compare the proposed DCF scheme against the standard DCF scheme.

We have plotted the normalized throughput for a node working proposed DCF mode (with CBR traffic) against standard DCF mode in Figure 6. With the proposed scheme, the mobile 6station is admitted to the BSS via admission controller and the scheduler schedules the allocated bandwidth for the station. Hence the normalized throughput stays close to one. With the standard DCF, the mobile stations share the bandwidth in uncontrolled manner. Thus, as the number of stations increases in the BSS, there is no control over the bandwidth usage.

SUMMARY

In this article we have proposed method of providing QoS in wireless LAN operating in DCF mode. Providing QoS guarantee in a distributed environment has to be a distributed approach due to the distributed nature of the DCF. The proposed method requires small changes (extensions) to the MAC management entity and scheduler function operating at each station above the MAC layer.

REFERENCES

Anker, T., Cohen, R., Dolev, D., & Singer, Y. (2001). Probabilistic fair queuing. In *IEEE Workshop on High Performance Switching and Routing.* Dallas, TX.

Bianchi, G. (2000). Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE JSAC,* 3, 535-47.

Chetan Kumar, S., Venkataram, P., & Pratap Singh, R. (2005). Distributed approach for QoS guarantee to wireless multimedia. In *International Conference on Advances in Mobile Multimedia,* Kuala Lumpur, Malaysia.

Network Simulator NS-2. (n.d.). Retrieved from http://www.isi.edu/nsnam/ns/.

Ni, Q., Romdhani, L., Turletti, T., & Aad, I. (2002). QoS issues and enhancements for IEEE 802.11 wireless LAN. In *Rapport de recherche de l INRIA.* Sophia Antipolis, Equipe: PLANETE.

Pong, D., & Moors, T. (2003). Call admission control for IEEE 802.11 contention access mechanism. *Globecom 2003,* San Francisco USA.

Tay, Y. C., & Chua, K.C. (2001). A capacity analysis for the IEEE 802.11 MAC protocol. *Journal of Wireless Networks, 7.*

KEY TERMS

Access Point (AP): An entity in the wireless LAN that is normally connected to a wired backbone and coordinates the operation of wireless mobile stations that are operating in infrastructure mode.

Call Admission Controller (CAC): A set of functions and procedures that control the admission of new calls into the system predictable based on predefined set of rules.

Media Access Controller (MAC): A set of procedures that governs the access to the media in a multiple access system. DCF (distributed coordination function) and PCF (point coordination function) are two MAC functions defined in IEEE 802.11 standard for wireless LAN.

Quality of Service (QoS): The term quality of service defines of quantitative representation of network resources that affect the application performance.

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 195-201, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.3
QoS Adaptation in Multimedia Multicast Conference Applications for E-Learning Services

Sérgio Deusdado
Instituto Politécnico de Bragança, Portugal

Paulo Carvalho
Universidade do Minho Braga, Portugal

ABSTRACT

The evolution of the World Wide Web service has incorporated new distributed multimedia conference applications, powering a new generation of e-learning development and allowing improved interactivity and prohuman relations. Groupware applications are increasingly representative in the Internet home applications market, however, the Quality of Service (QoS) provided by the network is still a limitation impairing their performance. Such applications have found, in multicast technology, an ally contributing for their efficient implementation and scalability. Additionally, considering QoS as a design goal at the application level becomes crucial for groupware development, enabling QoS productivity to applications. The applications' ability to adapt themselves dynamically according to the resources availability can be considered a quality factor.Tolerant real-time applications, such as video conferences, are in the frontline to benefit from QoS adaptation. However, not all include adaptive technology able to provide both end-system and network quality awareness. Adaptation, in these cases, can be achieved by introducing a multiplatform middleware layer responsible for tutoring the applications' resources (enabling adjudication or limitation) based on the available processing and networking capabilities. Congregating these technological contributions, an adaptive platform has been developed integrating public domain multicast tools, applied to a Web-based distance learning system. The system is user-centered (e-student), aiming at good pedagogical practices and proactive usability for multimedia and network

resources. The services provided, including QoS adapted interactive multimedia multicast conferences (MMC), are fully integrated and transparent to end-users. QoS adaptation, when treated systematically in tolerant real-time applications, denotes advantages in group scalability and QoS sustainability in heterogeneous and unpredictable environments such as the Internet.

INTRODUCTION

Technology has been a strong catalyst for educational innovation and improvement, especially when the World Wide Web is involved. The next generation Internet needs technological support to accommodate promising new applications, such as interactive real-time multimedia distribution. Predictable bandwidth availability and capacity solvency imply QoS management to regulate resources in heterogeneous environments. Actually, increasing the network capacity through advanced network and media technology is not *per se* a ubiquitous and definitive solution to overcome the network capacity problem. Historically, the users have always managed to consume the entire system capacity soon after it was enlarged (Ferguson & Huston, 1998). IP Multicasting techniques (Deering, 1998; Kosiur, 1998; Moshin, Wong, & Bhutt, 2001; Thaler & Handley, 2000) are attractive solutions for this capacity shortage problem as bandwidth consumption is reduced when network resources are shared. On the other hand, the QoS support (Moshin, Wong, & Bhutt, 2001) should be, in a first instance, inherent to applications in order to integrate conveniently enhanced real-time multimedia applications in the present Internet, barely QoS aware and increasingly heterogeneous.

With the advent of wireless and mobile networks, heterogeneity is likely to subsist; envisioned applications should merge QoS adaptation and multicast in a proactive utilization of resources. Applications should be designed with adaptation in mind; they need to employ built-in mechanisms that allow them to probe the conditions of the network environment and alter their transmission characteristics accordingly (Miras, 2002). Self-adaptive applications, in the sense of proactive behavior for transmission of continuous media in multiparty applications, are a well-accepted solution due to the correct integration of new services in today's Internet (Deusdado, 2002; Li, Xu, Naharstedt, & Liu, 1998).

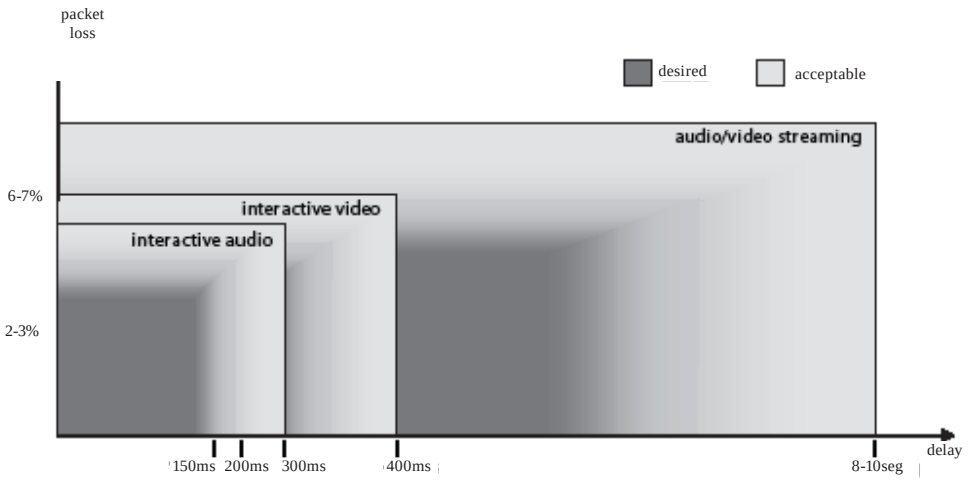
E-learning, as a component of flexible learning, encompasses a wide set of applications and processes which use available electronic media to deliver vocational education and training. It includes computer-based learning, Web-based learning, virtual classrooms and digital collaboration (Eklund, Kay, & Lunch, 2003). Our work aims to integrate interactive multimedia e-learning applications in a proactive fashion taking into account the available network resources and QoS sustainability. In this way, our motivation is to offer improved learning experience based on ultimate technology with QoS warranties.

The system architecture proposed in this article includes an adaptive module based on Java applets and embedded Javascript, responsible for assessing the existing operating conditions, by collecting metrics reflecting the client's end-system performance (e-student's host), the current network conditions and relevant multicast group characteristics. The collected data are subsequently computed weighting parameters such as the available bandwidth at the client side, the round-trip time between the client and the e-learning server, the client's current CPU load and free memory. The obtained results are used for proper multicast applications scheduling and parameterization in a transparent way.

MOTIVATIONS

Basically, e-learning services are used to promote connections between people (e-students)

Figure 1. QoS tolerance for generic audio and video applications (Miras, 2002)



and training resources (Steeples & Jones, 2002). E-learning research is wide and growing in importance, especially in higher education. Several institutions are developing interactive Web-based learning systems, integrating rich media streaming which may compromise network performance. The design of e-learning systems should consider QoS as mandatory for successful learning experiences, selecting the appropriate technologies and applications, and regulating proactively the information and communications technology (ICT) resources utilization (Allison, Ruddle, McKechan & Michaelson, 2001).

The Multicast Backbone (MBone) is a network overlaying the global Internet designed to support multipoint applications. MBone tools comprise a collection of audio, video and whiteboard applications that use Internet multicast protocols to enable multiway communications (point-to-multipoint and multipoint-to-multipoint), satisfying most of the needs of group communication, such as e-learning services. Using these applications by common e-students drives recurrently to poor QoS satisfaction due to the heterogeneity of resource conditions and the applications' inability to assess available conditions and adjust internal parameters before conference initiation. Without regulation, Real-time Transport Protocol (RTP) traffic floods the network capacity insensitively, forcing network congestion in certain cases or inhibiting better performance. A coherent behavior of an application without adaptation is difficult in today's Internet.

Public domain multicast applications used in this article, *vic* (McCanne & Jacobson, 1995), *rat* (Hardman, Kirstein, Sasse, Handley & Watson, 1995) and Java Media Framework (JMF) (JMF 2.0, 1999) were designed with no QoS "sensors," so the communication dynamics is not automatically interdependent of end-systems or network conditions. Effectively, such applications allow preparameterization to adjust critical parameters such as throughput, number of frames per second, video and audio encoding formats and so forth. Adaptation, in these cases, can be obtained by introducing a multiplatform middleware layer responsible for tutoring the applications' resources (adjudication or limitation) based on the available

processing and networking capabilities (Miras, 2002).

Common interactive real-time applications are fault-tolerant but suffer from QoS constraints; low-latency requirements and reliability are cumulative to achieve conference success. The diagram in Figure 1 attempts to illustrate the QoS tolerance, in terms of delay and packet loss, for generic interactive audio and video applications.

The main motivation of this article is to provide adaptive behavior to applications used on both sides of multimedia conferencing, focusing essentially on multicast members that initiate audio and/or video transmission. The underlying idea is to launch automatically MMC applications with proper audio and video codecs, bandwidth allocation inference and other parameters that affect sustainability and scalability during an e-learning session. Our emphasis is on the concept of "interactive e-learning services," relegating the concept of "e-learning course" to a secondary goal, which will be considered in future work.

Most prominent related work on friendly multimedia transmission over the Internet, based on a combination of system and network QoS feedback implementing equation-based adaptation is summarized in Bouras and Gkamas (2003) and Vandalore, Feng Jain, and Fahmy (2001).

SYSTEM'S ARCHITECTURE

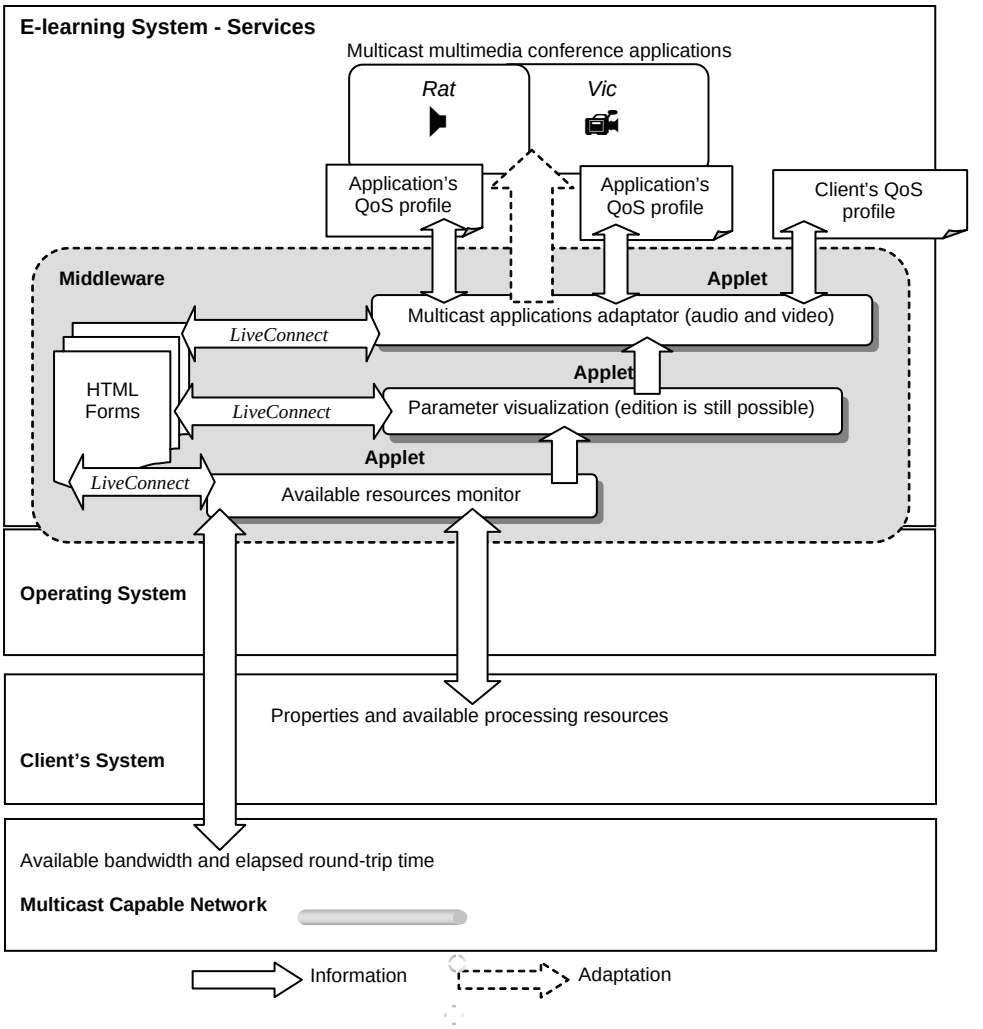
For multicast video distribution to heterogeneous users in an e-learning session, we assume that a class server (e-tutor's system) should be distributed and platform independent, considering inclusively multitutoring. Thus, a class server should connect to an e-learning server (Web server) and be submitted to adaptation as a regular new sender. The QoS requirements for the class server, operating in a centralized fashion, may justify the need of layered multicast (Johanson & Lie, 2002; Liu, Li & Zhang, 2004), enhancing the service's adaptation. However, this work aims at integrating e-students with heterogeneous equipment when they transmit audio and video to the group, as it happens in a conventional classroom. If a client (e-student) wants to interact and multicast video then the system's architecture will be integrated with fair adjustments attending both to the connection to the server (e-tutor) network and to the hardware processing capabilities. Client's adaptation should not depend on the other group members because they are transient, and consequently stability of transmissions could be very poor.

As the involved applications are characterized by an intensive use of host and network resources, the purpose of the middleware platform is to achieve by computation, on a scale of five differentiated modes, the proper integration of new multicasting members. Within this thematic, it means implementing an adaptive learners' participation in e-learning sessions by starting MMC applications transparently, with their functionality optimized for the current operating conditions.

To clarify these aspects Figure 2 illustrates the system architecture. As shown, three applets, operating sequentially and interdependently, are responsible for monitoring and assessing QoS conditions, inferring, announcing and/or editing computed adaptation parameters. The process culminates with the initiation of MMC applications, depending on the host and network profiles and covering eventual end-user explicit requirements.

Audio and video encoding formats, frame rate and other quality metrics may be chosen according to the resources' availability, providing coherent, friendly and fair participation in the network load balance. After monitoring sustainable network QoS with repeated measurements during approximately 15 seconds before media transmission, the Round-Trip Time (RTT) and bandwidth are calculated using a moving average. In addition, system's status variables, such as processor load, free memory, processor performance and so forth, are acquired, taking advantage of operating system facilities.

Figure 2. System's architecture



The system is multiplatform as the included applets differentiate the most popular operative systems (Windows and Unix), invoking appropriate inner services to obtain instant measures for the processor's load and free memory. The collected data constitutes another input to compute an adaptation index. Different compilations were produced for common browsers.

All of the adaptation process is transparent, however, regarding the experimental nature of this work, each phase allows interaction with the user, providing technical information or even accepting user preferences. To achieve this goal, applets and HTML forms interchange data using Sun's *Liveconnect* technology.

ADAPTIVE QoS FRAMEWORK

In the proposed framework, QoS management is performed individually for each new conference member and occurs before the transmission's start, such that MMC applications are launched adaptively facing the previous QoS sensing period

Table 1. vic and rat QoS parameters used to adjust applications' profile.

Rat	**-f** *format*	Indicates audio encoding format: *l16*, *pcm*, *dvi*, *gsm* and *lpc*
Vic	**-B** *kbps*	Sets the maximum bandwidth slider (kbps)
Vic	**-c** *dither*	On a color-mapped display, uses the algorithm indicated by dither (e.g., *ed*, *gray*, *od*, *quantize*) to convert to the available color palette
Vic	**-f** *format*	Indicates the video encoding format: *h261*, *h263*, *jpeg*, *nv*, ...
Vic	**-F** *fps*	Sets the maximum frame rate (fps)

Table 2. Set of parameters for different QoS adaptation modes

ADAPTATION MODE	MAXIMUM BANDWIDTH	FRAME RATE	VIDEO CODEC	COLOR	AUDIO CODEC
5	1 Mbps	30 fps	H.261	Yes	L16
4	512 kbps	25 fps	H.261	Yes	PCM
3	256 kbps	20 fps	H.261	Yes	DVI
2	128 kbps	15 fps	H.263	Yes	GSM
1	64 kbps	10 fps	H.263	No	LPC

conditions. QoS variability during the conference is not used to dynamically readapt the applications. If an e-student experiences lack of QoS while conferencing, the membership process should be restarted. Corroborating this practice, MMC applications, especially *vic*, are not stable enough. In fact, if some critical adjustments are made on-the-fly, the result is often the collapse of the application. Nevertheless, dynamic adaptation is currently a subject of study within group communication applications (Layaida & Hagimonte, 2002; Tusch, Böszörményi, Goldschmidt, Hellwagner & Schojer, 2004).

Considering the applications' specificity and type of traffic generated, adaptability only includes interactive audio *(rat*) and video *(vic)* applications and services. The heuristics regarding the choice of applications' QoS parameters emerged from experimental results and scientific references in this matter (Wu, Hou, Zhu, Zhang, & Peha, 2001). For instance, video conference users typically require better audio quality than video quality (Bolot, Crépin, & Garcia 1995). The success of video conferencing communication also depends on factors such as received frames per second, image quality, resolution, size and illumination.

For this work, the representative parameters of *vic* and *rat* used to modulate QoS are presented in Table 1.

The values for these parameters, deriving from a mathematical expression that generates an adaptation mode based on the sustainable QoS level, compose a set of adjusting directives determining the applications' behavior. Each adaptation mode indexes the respective set of adjustments, which will then be passed to the application. Since QoS scale varies from mode 1 to 5, when the obtained result is under or over this range it will be assigned to the nearest limit. Equation (1) determines the adaptation mode to be applied:

$$\mathbf{M = (int)\ (B/(RTT/2) + FM/P) *K} \qquad (1)$$

where [1],
M = QoS adaptation Mode (Table 2);
B = Bandwidth (kbps); **RTT** = Round-Trip Time (ms);
FM = Free Memory (MB); **P** = Processor load (%);
K = 1/50 - constant to scale the result (1 to 5).

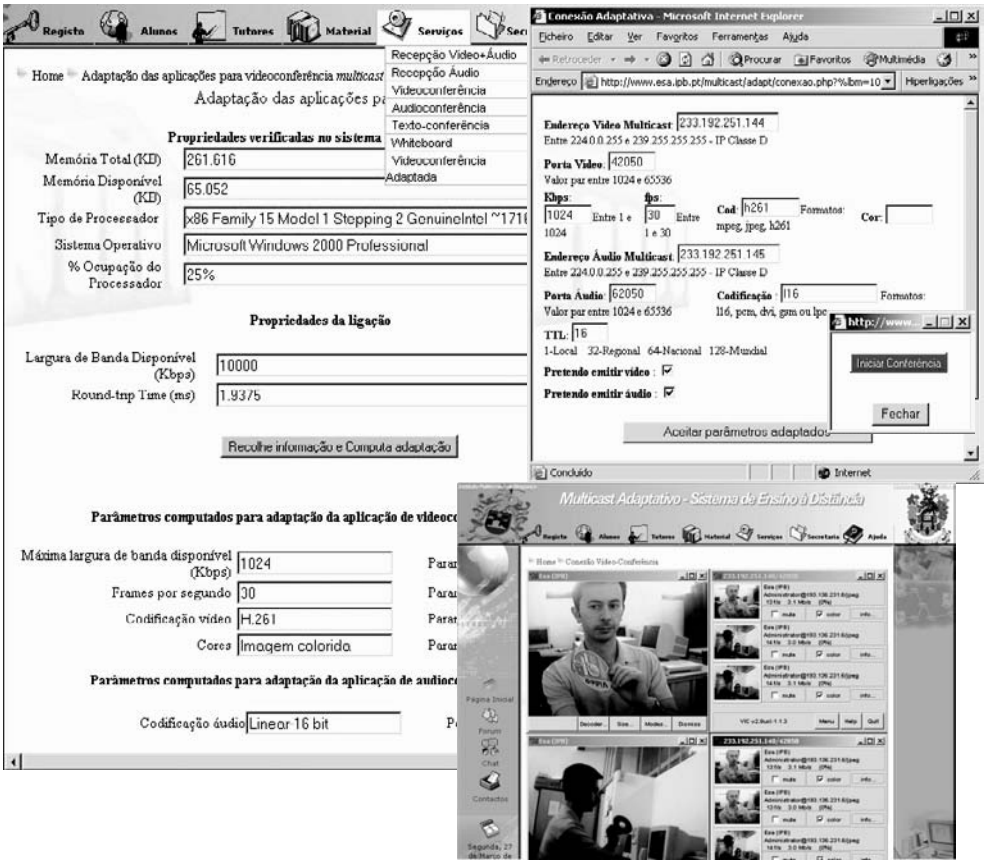
For *vic* (version 2.8), the video encoding formats H.261 (ITU-T H.261, 1993) and H.263 (ITU-T H.263, 1998) were those who revealed best performance for e-learning purposes, leading to low loss ratios and high reliability. H.263 is especially appropriate for low bandwidth environments.

E-LEARNING SERVICES AND FEATURES

The developed distance e-learning system presents numerous features providing distinct service levels, such as:

1. Virtual academy, Web-based with refined usability, integrating authentication and services for the e-learning community.
2. Registration, authentication and maintenance of educational agents.
3. Multicast sessions maintenance and scheduling.

Figure 3. E-learning system screenshots and MMC's adaptation HTML forms based on available QoS



4. Access to asynchronous material such as video on demand, slide presentations and other multimedia resources.
5. Interactive multimedia multicast conferences with QoS adaptation.
6. Other multicast tools for shared workspace.
7. Discussion spaces such as forum and multicast chat room.

This information system incorporates online databases structuring courses, students, tutors and sessions' data. These resources were developed using MySQL/PHP. A Web site congregating all developed application component prototypes is available at www.esa.ipb.pt/multicast.

Certain processes for assessing hardware performance require user's explicit authorization, allowing extended security privileges to applets in order to perform system's inspection and collecting substantial data used by subsequent applets of the control path. The security certificates used in this work are not provided by official entities, but generated by applet compilation tools for testing.

Although the adaptation process is totally transparent, effectively, the users may edit QoS parameters suggested by the system. If editing occurs, correctness and validation are assured by embedded *Javascript* code for parsing purposes.

Figure 4. BW needs for each QoS mode

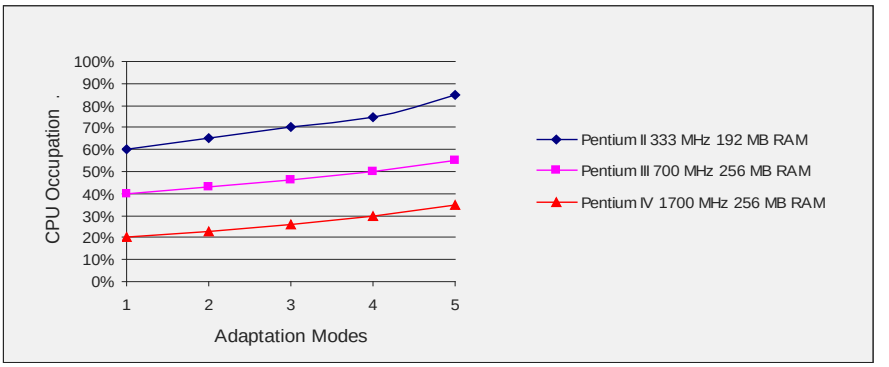


Figure 5. CPU needs for each QoS mode

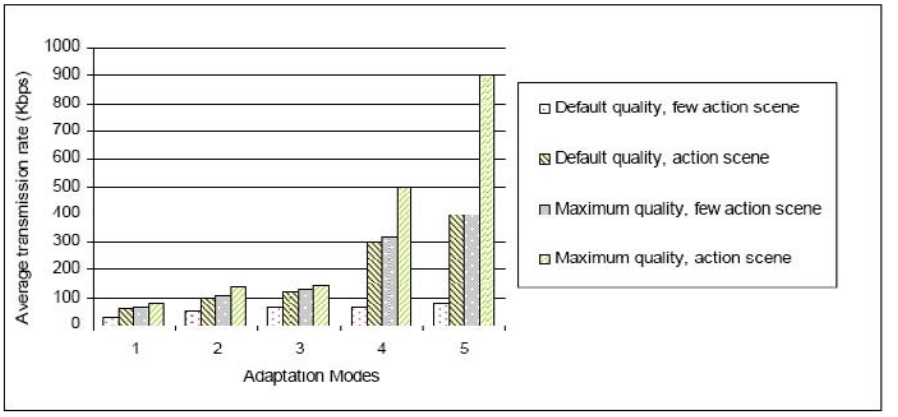


Figure 6 . Linear bandwidth distribution using applications' defaults, no adaptation is used.

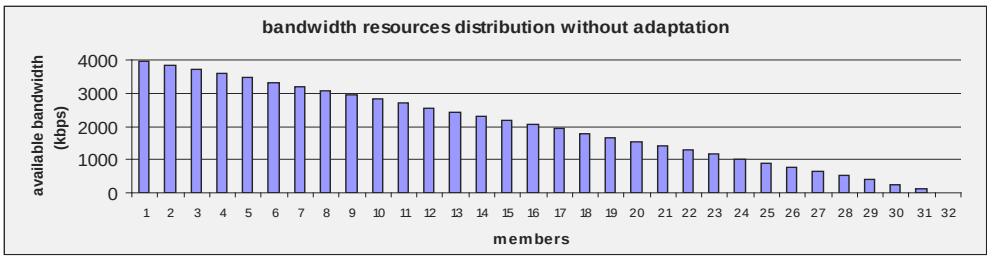


Figure 7. Increasing the number of active group members using adaptation to distribute network resources

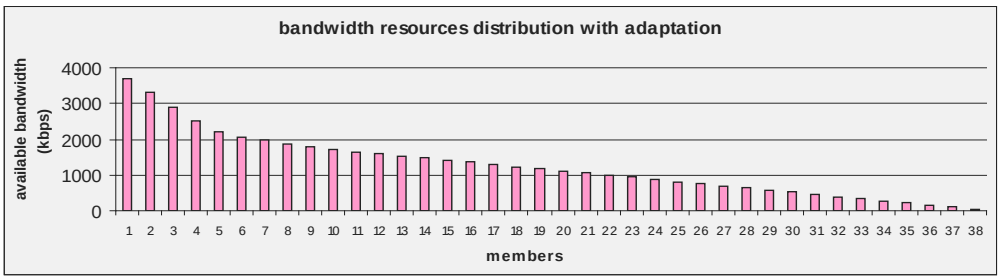
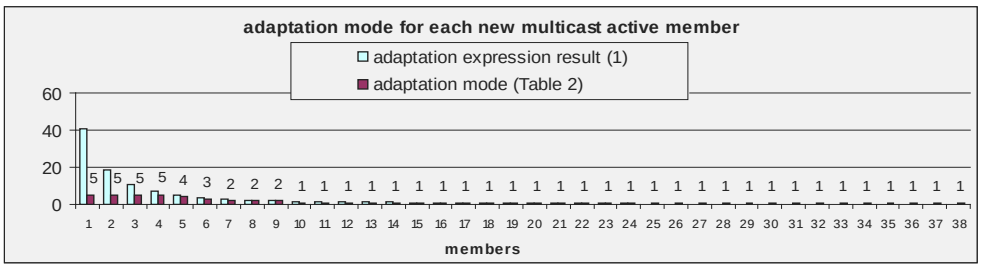


Figure 8. QoS mode adopted by the system facing the available resource conditions



All MMC applications need to be previously installed and accessible through the command line interface, configuring the PATH environment variable properly. If we want to transmit audio or video the required equipment must also be ready. Gathering these basic requirements it is possible to participate in e-learning sessions, having adapted QoS in a transparent way, with great usability. Figure 3 integrates a system's screenshots, illustrating step by step, when a QoS adapted video conference is selected from the "Services" menu.

PERFORMANCE STUDY

Applications that use voice, video streams or multimedia must be carefully managed within

an IP network to preserve their operational integrity. Beyond routing improvements, QoS in a multimedia conference needs primarily to deal with several sources with different characteristics, shifting large amount of traffic competing for network capacity. MMC applications may easily absorb all network resources and the subjective quality sensed by users, would remain poor if the available resources are used indiscriminately. As mentioned earlier, the adaptation purpose, with e-learning in mind, was to integrate MMC applications with QoS conscience, preserving resources in order to maintain conference quality and improve scalability.

In order to test the framework, different scenarios were simulated and the corresponding resources' consumption verified considering the QoS limitations associated with each QoS mode defined in Table 2. For video conferencing, regarding e-learning purposes, it is widely accepted that reference values correspond to "Maximum quality, few action scene." Bandwidth consumption in *vic* default mode is 128 kbps. When adaptation is requested, the different adaptation modes use the values charted in Figure 4. For instance, the best quality mode consumes around 400 kbps, allowing better image and motion.

Different equipment was also tested in order to validate the rank of the defined adaptation modes. We observed that modern high performance equipment tends to be neutral; in this case adaptation will be influenced overall by network conditions, but with mobile computation in mind, PDAs and cellular phones, CPU performance should not be relegated.

The experimental results were obtained varying the number of new multicast members transmitting voice and video, considering that all the multicast group members are multimedia receivers and transmitters capable using Any Source Multicast (ASM) technology via *vic* and *rat* applications. Due to the limited number of multicast monitoring tools publicly available, we use embedded applications resource meters and the Multicast Monitor (www.multicastmonitor.com) to collect and handle the resulting data.

The e-learning system, more concretely the tested prototype showed good performance indicators that validate the architecture model proposed. Because video traffic is quantitatively more representative of resource consumption, it was analyzed preferentially. Figures 4 and 5 exhibit the levels of resource consumption for each QoS mode considered. Here, the overhead introduced by middleware to prepare applications is marginal as it occurs before transmission time.

QoS adaptation, when treated systematically in tolerant real-time applications, denotes advantages in group scalability and QoS sustainability in heterogeneous and unpredictable environments such as the Internet and Mbone. Figures 6, 7 and 8 illustrate a comparison between two simulated sessions, the first without QoS adaptation and the second including adaptation managed by the developed middleware layer. The results show that scalability increased, but equally important is the fact that applications may benefit from resource availability that does not occur when using the default applications configuration. When the available resources decrease, the system allocates them to critical parameters. For instance, while the frame rate should not be below 10 fps, the image quality may be poor or monochromatic if the contents are correctly perceived.

Limiting bandwidth to applications, not only with explicit parameterization but also choosing the right encoding format for e-learning sessions, allows efficient resource utilization and proactive usability, avoiding network overload and congestion. If the network load remains high, it is easier to recover if adaptation is used.

The experience with Mbone showed that e-learning groups tend to be small, usually less than 20 members. Effectively, e-learning communities, as in traditional training methods, need a tutor, who is mainly an educational agent and not necessarily a learning technologist. Indeed, questions related with communication technologies will

not constitute pedagogical limitation if intelligent QoS management autonomy is provided natively or by middleware to applications.

E-learning conferencing specificity requires appropriate video encoding formats able to achieve low loss ratio and fast recovery from congestion. We compared H.261 and Motion JPEG (ITU-JPEG, 1992) performance in experimental sessions, using a modest PC (PIII 0.7GHz - 256 MB RAM). The results were penalizing for

MJPEG, where loss was about 30%, in opposition to 1% for H.261.

CONCLUSION AND DISCUSSION

The goal of this article was, in one topic, to foster "ecological" practices in the Internet when using MMC's applications in e-learning services. The proposed system integrates public domain multicast applications for synchronous media communication, being supervised by a middleware based QoS management framework, intending to preserve the QoS of critical parameters for e-learning session's specificity.

As main contributions, this work:

1. Provides an integrated e-learning environment based on interactive multimedia services with proactive QoS;
2. Improves the usability of MMC applications; and
3. Allows the development of end-to-end QoS-aware multimedia conferences, coordinating resources from network, end-system processing equipment and applications.

Middleware adaptation is a solution that suits the present state of Internet and the requirements of new multimedia distributed applications. We use a middleware layer to manage QoS adaptation in interactive audio and video applications coordinating resource demand, monitoring and adjudication. Substantive results were obtained in group scalability, QoS sustainability and proactive resource utilization.

Comprising multiple sources (even unauthorized ones), ASM involves high complexity and may compromise the success of e-learning conferences. Future work includes the use of Source-Specific Multicast (SSM) (Holbrook & Cain, 2004) in order to overcome this limitation. The development of new multilayer video encoding formats could also increase the flexibility when using QoS adaptation. When cumulative layers are transmitted avoiding redundancy, using different SSM groups or channels, adaptation can be performed in a transparent way in order to achieve efficient resources utilization (Johanson & Lie, 2002; Liu, Li & Zhang, 2004).

REFERENCES

Allison, C., Ruddle, A., McKechan, D., & Michaelson, R. (2001). The architecture of a framework for building distributed learning environments in advanced learning technologies. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, Madison,WI: IEEE Press.

Bolot, J., Crépin, H., & Garcia, A. (1995). *Analysis of audio packet loss in the Internet*. LNCS, 1018, pp. 154-165. France: INRIA.

Bouras, C., & Gkamas, A. (2003). Multimedia transmission with adaptive QoS based on real time protocols. *International Journal of Communications Systems*, *16*, 225-248

Deering, S. (1998). Multicast routing in internetworks and extended LANs. In *Proceedings of ACM SIGCOMM* (pp 55-64).

Deusdado, S. (2002). *Integração Adaptativa de Aplicações Multicast para Conferência Multimédia*. Unpublished master thesis, Universidade do Minho.

Eklund, J., Kay, M., & Lynch, H. (2003). *E-learning: Emerging issues and key trends.* Australian Flexible Learning Framework, Australian National Training Authority.

Ferguson, P., & Huston, G. (1998). *Delivering QoS on the Internet and in corporate network*s. New York: John Wiley & Sons.

Hardman, V., Kirstein, P., Sasse, A., Handley, M., & Watson, A. (1995). RAT, Robust Audio Tool. Retrieved July 27, 2006, from http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/

Holbrook, H., & Cain, B. (2004). *Source-specific multicast for IP.* Unpublished manuscript. Interent Draft, IETF, <draft-ietf-ssm-arch-07.txt>.

ITU-JPEG. (1992). JPEG Standard, *Information Technology- Digital Compression and Coding of Continuous-Tone Still Images- Requirements and Guidelines,* Recommendation T.81, ITU.

ITU-T H.261. (1993). ITU-T Recommendation H.261, *Video CODEC For Audiovisual Services At p x 64 kbits.*

ITU-T H.263. (1998). ITU-T Recommendation H.263, *Video Coding For Low Bitrate Communication.*

JMF 2.0 (1999). JMF- *Java ™ Media Framework API Guid*e, Sun Microsystems, JMF 2.0 FCS.

Johanson, M., & Lie, A. (2002). *Layered encoding and transmission of video in heterogeneous environments* (Tech. Rep.), Sweden: Department of Computer Engineering, Chalmers University of Technology.

Kosiur, D. (1998). *IP multicasting: The complete guide to interactive corporate network*s. New York: John Wiley & Sons.

Layaida, O., & Hagimonte, D. (2002). *Dynamic adaptation in distributed multimedia applications* (Tech. Rep.) INRIA Saint Ismier, France: Rhône-Alpes Research Unit,

Li, B., Xu, D., Naharstedt, K., & Liu, J. (1998). *End-to-end QoS support for adaptive applications over the Internet.* Unpubluished manuscript, University of Illinois at Urbana-Champaign.

Liu, J., Li, B., & Zhang, Y. (2004). An end-to-end adaptation protocol for layered video multicast using optimal rate allocation. *IEEE Transactions On Multimedia, 6*(1), 87-102.

McCanne, S., & Jacobson, V. (1995). vic: A flexible framework for packet video. In *Proceedings of the ACM Multimedia Conference*. Retrieved July 27, 2006, from http://www-mice.cs.ucl.ac.uk/multimedia/software/vic/

Miras, D. (2002). *A survey of network QoS needs of advanced Internet applications.* Unpublished manuscript, University College London.

Mohsin, M., Wong, W., & Bhutt, Y. (2001). *Support for real-time traffic in the Internet, and QoS issues.* Unpublished manuscript, University of Texas at Dallas.

Steeples, C., & Jones, C. (2002) *Networked learning: Perspectives and issues.* London, UK: Springer.

Thaler, D., & Handley, M. (2000). *On the aggregatability of multicast forwarding state*. Tel-Aviv, Israel: IEEE INFOCOM.

Tusch, R., Böszörményi, L., Goldschmidt, B., Hellwagner, H., & Schojer, P. (2004). Offensive and defensive adaptation in distributed multimedia systems. *ComSIS, 1*(1), 49-77.

Vandalore, B., Feng, W., Jain, R., & Fahmy, S. (2001). A survey of application layer techniques for adaptive streaming of multimedia. *Real Time Imaging, 7*(3), 221-235.

Wu, D., Hou, Y., Zhu, W., Zhang, Y., & Peha, J. (2001). Streaming video over the Internet: Approaches and directions. *IEEE Transactions on Circuits System Video Technology,11*(3) 282–300.

ENDNOTE

[1] **RTT** and/or **P** values will be, if necessary, assigned to 1 to avoid division by zero. To prevent incongruence, the maximum bandwidth allowed cannot exceed the detected value (**B**), otherwise the computed mode will suffer cyclic decrements while the excess remains and **M**>1.

This work was previously published in International Journal of Distance Education Technologies, Vol. 4, Issue 4, edited by S.-K. Chang & T. K. Shih, pp. 56-68, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.4
Quality of Service Issues in Mobile Multimedia Transmission

Nalin Sharda
Victoria University, Australia

ABSTRACT

The focus of this chapter is on the quality of service (QoS) aspects involved in transmitting multimedia information via mobile systems. Multimedia content and applications require sophisticated QoS protocols. These protocols need to manage throughput, delay, delay variance, error rate, and cost. How errors are handled in a multimedia session can have significant impact on the delay and delay variance. Coding and compression techniques also influence how the final presentation is transformed by the impediments encountered on a mobile network. Providing the user with the ability to negotiate between cost, quality, and temporal aspects is important, as this allows the user to strike a balance between these factors. In moving from 2G to 3G, and, over the next decade to 4G mobile networks, the ability to transmit multimedia information is going to improve constantly. Nonetheless, providers must develop viable economic models and user interfaces for providing differentiated QoS to the users.

INTRODUCTION

Transmission of multimedia information over mobile networks to portable devices, such as laptops, mobile phones, and PDAs (personal digital assistants), is leading to the development of new applications. However, successful transmission of multimedia information over mobile networks cannot be taken for granted. Understating the impediments to successful transmission of multimedia information is of paramount importance. This chapter focuses on multimedia applications that use mobile networks, and issues involved in the delivery of multimedia content with the desired quality of service (QoS). Current and future challenges in achieving successful mobile multimedia information transmission are also discussed.

Multimedia applications require more sophisticated QoS protocols than those for simple data transmission. The main parameters that underpin QoS are throughput, delay, delay variance, error rate, human perception of quality, and cost (Sharda, 1999). The interplay between these factors is rather complex, therefore, some simplifying assumptions must be made in developing

methodologies for delivering multimedia content with the desired QoS.

For the delivery of desired QoS, one of the most promising concepts developed over the last few years is that of resource reservation. This entails reserving resources such as bandwidth on interconnects, and buffer space and processing power on switching nodes.

Packet switching networks embody the idea of statistical time division multiplexing (STDM); that is, resources are allocated to a communication session based on the demands of the traffic. This leads to more efficient, and therefore, more economical usage of the resources. However, the need to allocate resources dynamically adds complexity to the communication system's operation and management. Mobile multimedia communications are further complicated due to their variable transmission quality, the need to keep track of end system location, restrictions placed due to limited battery life, reduced screen size, and the cost of the connection.

Over the last decade, some progress has been made in establishing mobile multimedia transmission systems. However, much research and development is still required before we can take it for granted that a multimedia application, such as videoconferencing, would run with the desired QoS over a mobile communication infrastructure on a hand-held device as we zoom down a freeway at high speed, and, all this at a reasonable cost.

The next section of this chapter presents the challenges introduced by the mobile multimedia content, applications, and communication systems. It begins with an overview of mobile multimedia systems, and then presents the implications of coding and compression techniques for transmitting multimedia. Requirements of various multimedia applications and their relationship to mobile communication systems are also presented.

The third section presents QoS issues in transmitting multimedia content over mobile systems. Fundamentals of QoS concepts and different QoS models are introduced, and a novel model for managing QoS in real time is presented.

The fourth section presents directions for future research, and the final section gives the conclusions.

MOBILE MULTIMEDIA SYSTEMS

Overview

This section presents an overview of coding methods used for various media types, multimedia applications, and current mobile communication systems. QoS issues related to each of these are also discussed.

Multimedia communication systems combine different types of media contents, such as text, audio, still images, and moving images, to achieve the overall objective of a communication session. Therefore, the network needs to provide a service which works well for all media types.

The requirements for successfully transmitting a particular media type depend upon its coding and compression techniques, and the application in which it is being used. Media content that must be transmitted live, or processed in real time, poses more stringent requirements. Consequently, live video conferencing is one of the most challenging multimedia applications.

The network infrastructure and the communications protocols used for transmission play a vital role in satisfying the demands of a given application. In general, multimedia transmission requires high bandwidth, low error rate, low delay, and very low delay variance. To date, we have not solved all of these challenges for even wired media. Fulfilling these requirements for achieving high-quality multimedia transmission over wireless connections is even more challenging.

The transition from the 2nd generation (2G) mobile systems to the 3rd generation (3G) mobile communication infrastructure presents new opportunities; however, still there are many chal-

lenging problems that need to be overcome. One of the key features missing in the current systems is the facility for the user to negotiate with the system and strike a compromise between the three key service aspects—quality, cost, and time—just as any market-oriented goods or services have to strike a balance between the quality, cost, and its delivery time.

Errors encountered in any transmission system can be either ignored, or detected and corrected. Errors can be ignored only if the received message is usable even with some errors. If errors in the received message are not acceptable, then these errors must be detected and corrected. Reverse error correction protocol requests retransmission of packets received with errors. This not only adds delay to the final reception of packets, it also adds delay jitter, as different packets encounter different delays. Forward error correction protocols include additional error correction bits, so that some of the errors can be corrected from the received data; this adds to the total data traffic. The choice of error handling method depends upon the type of data, its coding methodology, and the application.

Multimedia Content

By definition, a multimedia system combines different media types: text, audio, still and moving images. Each of these content types can be further categorised into sub-types. For example, still images can be bi-tonal, greyscale, or full-colour; furthermore, these can have continuous variation in tone—as in a photograph, or have sudden variation in the intensity—as in a printed page. A variety of techniques are used for digitally coding still images, depending upon the image type and application. Similarly, many text representation techniques and associated digital coding techniques are used. Audio and video are even more complex, as these are time varying quantities and involve continuous sampling over time. Errors and delays introduced at any stage of sampling, encoding, transmitting, and decoding of audio and video can lead to reduction in the quality of the final presentation.

Most multimedia content needs to be compressed to reduce the storage space and transmission bandwidth. Uncompressed multimedia content has in-built redundancy, and a few corrupted bits do not change the contents dramatically. Conversely, compressed media is compact, and has much less redundancy. Consequently, any errors during transmission affect compressed content more severely.

Mobile transmission systems are inherently more error-prone than wired transmission systems. The requirements for successfully transmitting a particular media type over a network depend not only on its coding and compression techniques, but on its application as well. However, all multimedia content is for human consumption, therefore, the criteria for acceptable quality of presentation ultimately depends upon human perception. For example, streamed video can accept a few seconds of delay, but live video conferencing becomes rather ineffective if the round-trip delay exceeds even a tenth of a second.

Text Coding

Despite the move towards graphical information, text remains a vital part of any multimedia presentation. One of the most enduring text codes is the American Standard Code for Information Interchange (ASCII). ASCII began its life as a 7-bit code designed for use with teletypes. Today, if someone talks of an ASCII document, they essentially refer to a text document with no formatting. Applications such as Notepad create ASCII text, and word processors can save a file as "text only". Extended ASCII codes were designed for computers to be able to handle additional characters from other languages. It took some time to get a single standard for these additional characters, and there are a few Extended ASCII sets.

Unicode provides a text code that is independent of platform, program, or language. In Unicode, a unique 16-bit number is reserved for every character. The Unicode standard aims to provide a universal repertoire with logical ordering that is efficient. The latest version of Unicode Standard is Unicode 4.0.1, and supports around a hundred international scripts.

ASCII and Unicode have been used extensively over wire-line communication systems, and can be used over wireless media as well. In general, transmission of text codes does not require high bandwidth or stringent limits on delay and delay variance. Hence maintaining QoS in transmitting text is often not much of a problem. Nonetheless, a new code set was designed for sending short text messages over mobile systems.

Short message service (SMS) uses a 7-bit code set that enables one to send and receive text messages of up to 160 characters on mobile phones. Some 8-bit messages are used for sending smart messages (such as images and ring tones) and for changing protocol settings. For Unicode-based text messages, 16-bit codes of maximum 70 characters can be used. These are viewable by most phones, and some appear as a flash SMS, that is, appear on the screen immediately upon arrival, without pressing any button. The SMS code was originally developed for the 2G technology, and therefore works well with 2G as well as 3G systems. The only possible issue with respect to QoS can be errors; bandwidth, delay, and delay jitter do not impede the transmission of SMS messages.

Non-Textual Information

The standard developed to transmit multimedia information over the Internet is the multipurpose Internet mail extensions (MIME). This standard was developed by the Internet Engineering Task Force (IETF) to support the transmission of mixed-media messages across TCP/IP networks. This also became the standard for transmitting foreign language text which the ASCII code could not represent.

Multimedia messaging service (MMS) provides the ability to send messages that combine text, sounds, images, and video over wireless networks. This requires handsets that are MMS capable. MMS is an open wireless standard specified by the WAP (wireless application protocol) forum—which has now been consolidated into the Open Mobile Alliance (OMA). In the WAP protocol, a notification message triggers the receiving terminal to start retrieving the message automatically using the WAP GET command. This retrieval may be modified by applying filters defined by the user. The content that can be transmitted with the WAP protocol can use a variety of media types and encoding standards.

Audio Coding

The basic technique for digitising analog audio signals is called pulse code modulation (PCM). In this technique, an analog audio signal is sampled at a rate double that of the maximum frequency that needs to be captured, and each sample is stored using 8-bit or 16-bit words.

Phone quality audio signals are sampled at 8,000 samples per second, and stored with 8-bit resolution; this generates 64 Kbps data rate. CD quality audio has two channels; it is sampled at 44,000 samples per second and saved with 16-bit resolution, giving a 1.4 Mbps data rate. A variety of compression techniques are used to reduce the bandwidth required to transmit audio signals. Compression becomes particularly important for CD quality stereo music, as the required 1.4 Mbps bandwidth is not economically available even in wire-line networks, much less so in wireless networks.

The MP3 (MPEG audio Layer 3) compression format has become one of the most widely used standards for transmitting high quality stereo audio. MP3 is one of three audio coding schemes associated with the MPEG video compression

standard. The MP3 standard provides the highest level of compression and uses perceptual audio coding and psychoacoustic compression to remove all redundant and irrelevant parts of a sound signal that the human ear does not hear. MP3 uses modified discrete cosine transform (MDCT) and improves the frequency resolution 18 times with respect to that of the MPEG audio Layer 2 coding scheme. It manages to reduce the CD bit rate of 1.4 Mbps down to 112-128 Kbps (a factor of 12) without sacrificing sound quality. Since MP3 files are small, they are easily transferred across the Internet, and are also suitable for transmission over wireless networks.

The next generation of MP3 standard is called mp3PRO. It is fully compatible with MP3, while halving the storage and bandwidth requirements. With this standard CD quality stereo can be transmitted at 64 Kbps. Furthermore, it can be used with digital rights management software, and can be ported transparently to any MP3-friendly application.

Advanced audio coding (AAC) is a wideband audio coding algorithm that exploits two main coding strategies to reduce the amount of data needed to encode high-quality digital audio. First, it removes signal components that are not important from a human perception point of view, and second, it eliminates redundancies in the coded audio signal. The MPEG-4 AAC standard incorporates MPEG-2 AAC, for data rates above 32 Kbps per channel. Additional techniques increase the effectiveness of the AAC technique at lower bit rates, and are able to add scalability and/or error resilience. (These techniques extend AAC into its MPEG-4 version: ISO/IEC 14496-3, Subpart 4.) The MPEG-4 aacPlus standard combines advanced audio coding techniques such as spectral band replication (SBR), and parametric stereo (PS). The SBR techniques deliver the same audio quality at half the bit rate, while the PS techniques (optimised for the 16-40 Kbps range) provide high audio quality at bit rates as low as 24 Kbps (Dietz & Meltzer, 2002).

The aacPlus codec family includes two versions. Version 2 of aacPlus is the high quality audio codec targeted for use in the 3GPP (3rd Generation Partnership Project). The aacPlus version 1 standard is adopted by 3GPP2 and ISMA (Internet Streaming Media Alliance) for digital video broadcasting (DVB).

The relationship between the aacPlus codec family members is shown in Figure 1 (Dietz & Meltzer, 2002). To compress the incoming stereo audio, the encoder extracts parametric representation of the stereo aspect of the audio. The stereo parametric information takes 2-3 Kbps and is transmitted along with the mono signal. Based on the parametric representation of the stereo information, the decoder regenerates the stereo signal from the received mono audio signal.

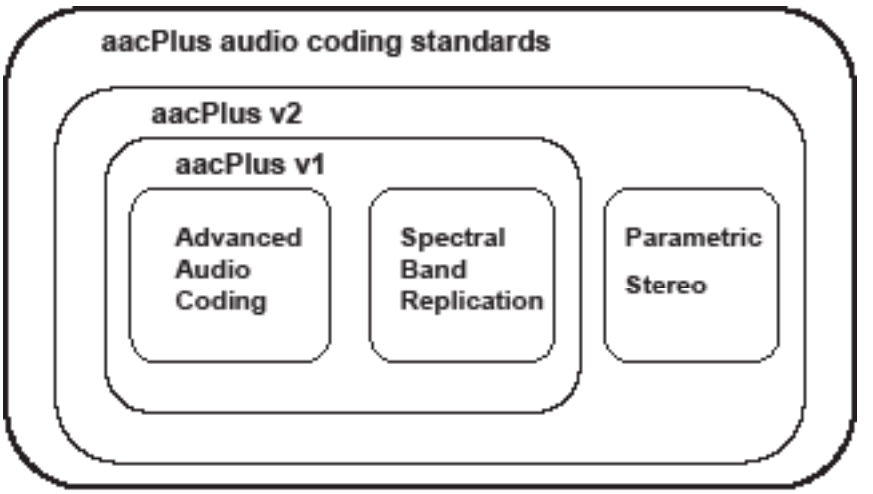
To be able to transmit high quality stereo audio, it is necessary to compress it to reduce the bandwidth, otherwise it may not be possible to obtain the desired QoS, especially over wireless networks. However, high level of compression makes the transmitted signal highly susceptible to errors, especially if the audio is being transmitted in real time. Any loss in the parametric information will severely degrade the quality of the reproduced stereo signal. Stereo is often used for music, and the slightest imperfection in music gets noticed by even non-experts.

Human ears are more sensitive to errors than human eyes. Human hearing faculties behave like differentiators, accentuating any variations, while human eyes behave like integrators, smoothening out variations (Sharda, 1999). Therefore, an audio stream should be given higher priority as compared to text or an image data stream.

Still Image Coding

Still image coding depends upon the type of image and its compression algorithm. Standards such as JPEG (Joint Photographic Experts Group), GIF (Graphics Interchange Format), and PNG (Portable Network Graphics) have dominated the field so

Figure 1. Relationship between aacPlus audio codecs v1 and v2 (Dietz & Meltzer, 2002)



far. JPEG is generally used for lossy compression of continuous tone images, such as photographs. GIF is a bitmap image format for pictures with 256 colours. PNG is a lossless bitmap image format. PNG improves upon the GIF format and is freely available.

The newer JPEG 2000 image compression standard uses a wavelet transform instead of the discrete cosine transform used in JPEG (Taubman & Marcellin, 2002). Therefore, JPEG 2000 can give higher compression ratio without generating the blocky and blurry artefacts introduced by the original JPEG standard. It also allows progressive downloads to extract various image resolutions, qualities, components, or spatial regions, without having to decompress the entire image. Distortion performance is also improved over the original JPEG standard, especially at low bit rates and at extremely high quality settings. JPEG 2000 is more error resilient as compared to the original JPEG standard (Secker & Taubman, 2004). This makes JPEG 2000 much better suited for applications requiring image transmission over wireless networks, as errors and delays introduce fewer observable artefacts in the displayed image.

Wireless networks experience higher error rates, and have lower bandwidth. Therefore, they are more severely challenged when transmitting digital images. Since the JPEG 2000 standard provides higher compression ratios, it is more suitable for the low bandwidth wireless networks; however, some additional issues need to be addressed (Santa-Cruz, Grosbois, & Ebrahimi, 2002). Issues such as error resilience over wireless networks are being addressed by the JPEG 2000 Wireless (JPWL) team. Their aim is to standardise tools and methods for efficient transmission of JPEG 2000 images over error-prone wireless networks. One of the techniques being developed by JPWL make the code stream more error resilient by adding redundancy, or by interleaving data (Dufaux & Nicholson, 2004). The decoder not only detects errors, but also corrects some, where possible. Another technique changes the sensitivity of different parts of the code stream to errors. More sensitive sections of the code stream are more heavily protected than the less sensitive sections. The third technique describes the locations of the remaining errors in the code stream;

the decoder then uses this information to exclude the corrupted parts of the code stream from the decoding process.

By standardising these techniques in JPWL, JPEG 2000 is being made more resilient to transmission errors, making it an ideal choice for the transmission of digital images and video over wireless media.

Moving Image Coding

Moving image coding can entail storing up to 20-30 image frames in every second. This demands very high bandwidth for high quality uncompressed video. Development of video coding standards that provide low resolution and low frame rate video suitable for transmission over networks began with the H.261 standard published by the ITU (International Telecom Union) in 1990, with data rates in multiples of 64 Kbps. The H.263 version provided a replacement for H.261 (in 1995) to work at all bit rates. It was further enhanced as H.263v2 (in 1998) and H.263v3 (in 2000). H.263 is similar to H.261, with improved performance and error recovery, and supports CIF,[1] QCIF, SQCIF, 4CIF, and 16CIF images. As these standards are designed for multiples of 64Kbps rates these are sometimes called px64 (where p can be 1-30). Originally these data rates were expected to suit ISDN (integrated services digital network) lines, nonetheless, these standards are useful in transmitting video over other wire-line and wireless networks also. H.263 is the baseline standard for the new 3G-324M standard, which targets the 3G wireless networks (Smith & Jabri, 2004).

Another option within the 3G-324M specification is the next generation video coding standard MPEG-4 AVC. It was approved in 2003 and called MPEG-4 AVC or ITU-T H.264, or simply advanced video coding (AVC).

MPEG-4 AVC doubles the compression efficiency of earlier standards for the same picture quality, which leads to 50% lower bandwidth (Navakitkanok & Aramvith, 2004). Therefore, it is far better than the earlier standards for wireless transmission. It offers improved resilience to transport errors, improved bit rate scalability, and stream switching for transmission over less reliable network infrastructure, such as wireless networks.

Motion JPEG 2000 (like Motion JPEG) can perform video compression applying only intra-frame compression. This makes Motion JPEG 2000 well suited for video transmission over wireless networks. It has been demonstrated that Motion JPEG 2000 outperforms MPEG-4 in terms of coding efficiency, error resilience, complexity, scalability, and coding delay (Tabesh, Bilgin, Krishnan, & Marcellin, 2005).

The JPWL work has taken into consideration the general principle underpinning networking protocols, with particular attention given to 3G networks (3GPP/3GPP2), wireless LANs (WLAN based on the IEEE 802.11 standards family), and Digital Radio Mondiale (DRM), making motion JPEG 2000 particularly suitable for wireless networks.

Multimedia Applications

Applications which have so far been bound to wire-line networks and desktop computers, now want to be let loose. The only option is to use wireless networks and portable devices. Areas for such mobile multimedia applications include both personal and business communications. E-learning, marketing, travel, and tourism are just a few of the burgeoning application areas that can make good use of mobile multimedia systems. Some of the potential killer applications based on the JPEG 2000 Wireless (JPWL) methods include video streaming and video conferencing (Liu & Choudary, 2004).

Mobile systems offer new opportunities and challenges as they become capable of transmitting multimedia information. Such applications need to transmit not only the core information, but also some associated meta-information. Most

electronic systems use multi-tier information transmission processes, which include: intimation of arrival (bell, ring, beep, and vibrate); abbreviated information (subject, caller ID); textual information (text message, SMS); multimedia information, and meta-information for layered retrieval of the information.

How a particular information type and associated meta-information is used depends upon the application, the user preference, user device, and the required QoS (Cheng & Shang, 2005).

Text Applications

Text is very useful for communication. It is often said that a picture is worth a thousand words; nonetheless, we should not forget that a few well chosen words can be worth scores of pictures. Additionally, text requires much lower bandwidth, and has greater certainty of meaning. It is more reliable in the face of transmission errors, especially if we use either reverse or forward error correction protocols.

Text is easy to transmit asynchronously or synchronously. One can send an SMS to a friend during work, without the fear of disturbing her in an import meeting. The receiver can reply in her own time, or the two can engage in a brief chat session to fix their evening rendezvous. The runaway success of SMS follows the predicate that "brevity is the soul of wit," as SMS allows succinct messages that convey the meaning quickly. Coded messages based on SMS have also become prevalent, further reducing the time taken to enter and read the message.

Some commonly used SMS codes include: *ATB—All the best; BRB—Be right back; GR8—Great; LUV—Love; PCM—Please call me; TTYL—Talk to you later; 2DAY—Today; and WER R U—Where are you?* SMS codes are also being used to download information to mobile phones, such as snow photos to check the condition on ski slopes.

In Japan, codes called Emoji have been developed. These are colourful, and often animated inline graphics used for mobile messaging. However, these are not standardised or interoperable between carriers. Emonji's are treated as characters, and each carrier has its own set.

In conclusion, text or text-like messages are, and will remain, an important aspect of mobile communications, especially because these are inexpensive, highly expressive, and are least problematic with respect to delivery with the desired QoS.

Audio Applications

Transmission of voice was the original motivation for developing the mobile communications technology. However, digital radio is also coming online, and integrating digital radio in mobile phones is in the offing. An Austrian company Livetunes has developed UMTS-enabled handset with digital radio. SIRIUS Satellite Radio can transmit commercial-free music and other audio entertainment to cars and homes. Mobile audio commercials over such digital radio channels allow advertisers to send audio commercials to their customers' mobile phone. The customer receives a phone call; upon answering the call, the audio commercial is played. It can include new offers, promotions, and announcements. To avoid spamming, companies have to provide their own subscriber database and the audio clip.

The QoS requirements for audio are different for bi-directional conversation than those for uni-directional digital radio transmission. For digital radio, buffering can be used to remove any delay jitter; however, excessive buffering can add unacceptable delay to conversational applications (Sharda, 1999).

Human hearing is very sensitive to any distortion in audio. For conversational audio, we can tolerate some errors, as long as the meaning of the spoken words is clear. If there is a problem in understanding the meaning, then the listener can

always ask the speaker to repeat what was said. This is like reverse error correction working at the highest communication layer, that is, the user layer. However, this cannot work for stereo music; as human hearing works like a differentiator, and any distortion gets accentuated. Furthermore, our hearing is capable of picking slightest variation between the two channels of stereo music. Therefore, QoS issues are very important when high quality stereo music is transmitted, but not so important for conversational audio.

Still Image Applications

Applications needing still image transmission can use multimedia messaging service (MMS). Examples of MMS based applications include: weather reports giving images, stock prices displayed as graphs, football goals displayed as a slide show, and many more. An extension of still image transmission is animated text messages. The main QoS factor that affects still images is delay. As delay jitter does not effect still image transmission, any errors can be overcome by using reverse error correction protocols. If such additional delays are not acceptable, then images can be displayed with errors. Uncompressed images can tolerate a high level of errors; however, the ability to tolerate errors reduces for compressed images. The original JPEG type compression techniques lead to blocky images when errors occur, as they use discrete cosine transform. However, JPEG 2000 compression standard overcomes this problem by using wavelet transform. Images compressed with JPEG 2000 degrade "gracefully" in face of errors. As wireless communication systems are inherently more error prone, image-based applications will benefit from the use of JPEG 2000 standard for their compression (Dufaux & Nicholson, 2004).

Content repurposing is also becoming important, so that the content creator can compile content only once, and the system can vary image size and resolution depending upon the display screen size and the communications channel bandwidth (Rokou & Rokos, 2004)

The aim of content repurposing is to push the content with the most appropriate resolution, so that it can be transmitted over the available network to meet the QoS goals. In general, this would imply pushing lower resolution images over wireless networks. However, with JPEG 2000 and JPWL, the system can push a rough image to begin with, which keeps improving as more data bits are transmitted.

Video Applications

Video phones are a natural extension of the current audio telephony. Wire-line based video phones were demonstrated decades ago, however, these never became popular. Mobile video telephony is likely to become popular once the cost of transmitting acceptable quality video becomes affordable. In the meanwhile, look-at-this (LAT) applications will generate demand for mobile Internet and 3G wireless networks, as these will create large amount of real-time mobile video. Some possible LAT application areas include:

a. **Retail:** Before purchasing an item, the consumer sends an image of the item to their partner for comment or approval.
b. **Real Estate:** An agent sends images of the building and its surrounding areas to the prospective customer.
c. **General Business:** A worker sends live video to colleague(s) at other location(s) while holding a voice conversation. This could be applied to developing new ideas; designing new products; repairing faulty equipment; maintaining, installing, or inspecting a system.

QoS is of great importance in video transmission. Video conferencing is the most challenging multimedia application for transmission over mobile systems. Much effort has gone in to migrating

from 2G networks to 3G networks to provide the desired QoS for video transmission. However, the cost of transmission is still high enough for it to be an impediment in its large-scale adoption.

Mobile Communication Systems

The desire to communicate over long distances has been an innate need for human beings since time immemorial. We can reflect that the earliest telecommunications systems devised by human beings were wireless systems, namely, smoke signals, semaphore flags, drums, and yodelling and so forth. Therefore, it is not surprising that electronic communications are also moving towards wireless systems.

Evolution of Telecommunications

Electric telecommunications began with the telegraph demonstrated by Morse in 1837 and the telephone developed by Bell in 1876. Marconi began his experiment with radio transmission in 1895. Automation of circuit switching systems began in 1919 with the Strowger exchange. The era of satellite communications dawned with the Telstar satellite in 1950. Saber became the first major data network in 1962.

Evolution of Mobile Systems

An early landmark in the development of wireless communications was the patent for the spread spectrum concept, proposed in 1941 by Hedy Lamarr. The first mobile telephone service was setup in St. Louis by AT&T as far back as 1946. Some theoretical breakthroughs also occurred around this time. In 1948, Claude Shannon published the Shannon-Hartley equation, and in 1949 Claude Shannon and Robert Pierce develop the underlying concepts for CDMA (code-division multiple access). In 1950, Sture Lauhrén made the world's first cellphone call, and by 1956, Swedish PTT Televerket operated a mobile telephone service. In 1969, the Nordic Mobile Telephone Group started a mobile service. CDMA was deployed for military systems in the 1970s. In 1973, Motorola vice presidents Marty Cooper and John Mitchell demonstrated the first public call from a handheld wireless phone.

Evolution of Digital Mobile Systems

First Global System for Mobile Communications (GSM) technology based networks were implemented by Radiolinja in Finland in 1991. In 1992, the Japanese Digital Cellular (JDC) system was introduced. By 1993, the IS-95 CDMA standard got finalised. First meetings of the 3GPP (3rd Generation Partnership Project) Technical Specification Group was held in December 1998. In 2000, Siemens demonstrated the world's first 3G/UMTS (3rd Generation Universal Mobile Telecommunications System) call over a TD-CDMA (time division-CDMA) network.

In 2000, commercial GPRS (general packet radio service) networks were launched. These networks supported data rates up to 115 Kbps, as compared to GSM systems with 9.6 Kbps data rates. In 2001 NTT (Nippon Telegraph & Telephone Corp.) produced commercial WCDMA (wide-band CDMA) 3G mobile network. In 2003, Ericsson demonstrated the transmission of IPv6 traffic over 3G UMTS using WCDMA technology.

Fixed Wireless vs. Mobile Communications

We need to distinguish between fixed wireless communication systems and mobile communication systems. A mobile communication system frees the end systems from the tyranny of being connected to a wall socket, and provides the ability to communicate anytime and anywhere. It allows the freedom to roam outside the home or the office.

Fixed wireless communication systems are local alternatives to wired communication systems. These do not provide mobility outside the home or the office, nonetheless, they provide a cost effective telecommunications connection for a given location with the ability to move around within a specified boundary.

For remote locations, satellite-based communication systems may be the only means of establishing a connection; however, these can be expensive. Satellite connections add about half a second round trip delay, making full-duplex audio or video connections rather difficult. While there is appreciable delay in a satellite connection, the delay variance is not very high, as the number of hops is fixed at two—transmitter to the satellite, and satellite to the transmitter. Error rates can be high on a satellite connection, especially burst errors—in case of atmospheric disturbances.

Universal Mobile Telecommunications System (UMTS)

Today, there are more than 60 3G/UMTS networks using WCDMA technology. Over 25 countries have adopted this technology, and there is a choice of over 100 terminal designs in Asia, Europe, and the U.S. The 3G mobile technologies identified by ITU for 3G/UMTS offer broadband capabilities to support a large number of voice and data customers, and offer much higher data rates at a lower incremental cost than the 2G technologies (Myers, 2004).

One of the issues driving the development and proliferation of 3G technologies is the recognition that there is a need for guaranteeing the QoS for multimedia traffic. Without guaranteed QoS, many applications fail to perform as per the users' expectations. Until the users are confident of getting the quality they need for running mobile multimedia applications effectively, they will not shift from their current mode of operation and adopt the new wireless networking technologies for multimedia information transmission.

QUALITY OF SERVICE IN MOBILE SYSTEMS

This section gives an overview of the various approaches being trailed for the provision of QoS in mobile networks. While much work has been done in providing QoS guarantees at the network infrastructure level, a holistic approach to providing end-to-end QoS has been missing to some extent. We begin by presenting a QoS model that focuses on the user, and develops a methodology for allowing the user to negotiate with the system to find a compromise between cost, quality, and temporal aspects.

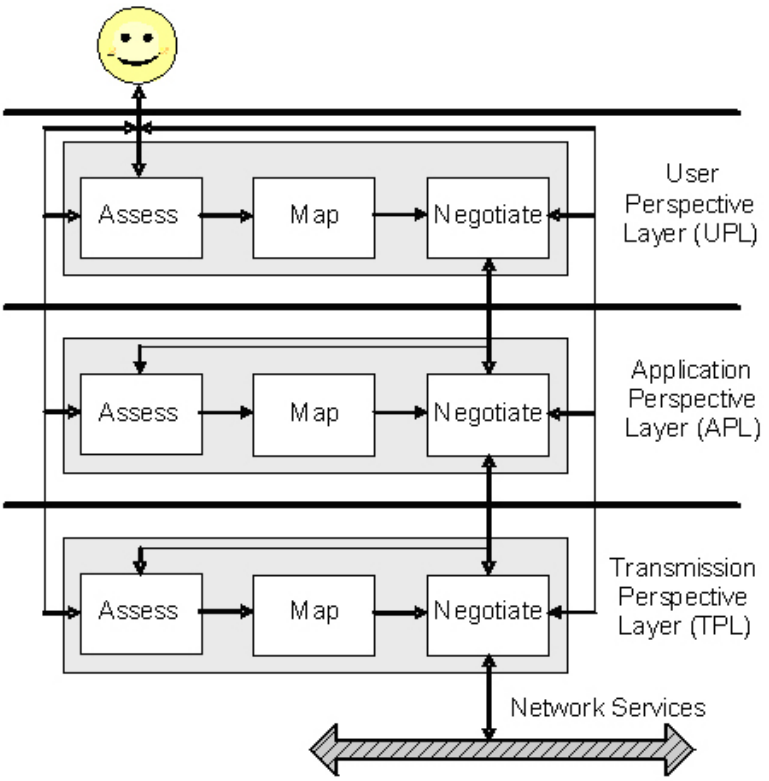
QoS Concepts and Models

The three layer quality of service (TRAQS) model shown in Figure 2 comprises three layers for QoS management in multimedia communications (Sharda & Georgievski, 2002). These three layers are: user perspective layer, application perspective layer, and transmission perspective layer. Each layer performs QoS processing for a set of QoS parameters that are related to the specific perspective. The main functions of the three perspective layers are:

- **User Perspective Layer (UPL)** interacts and performs QoS negotiations with the user and then transfer the QoS request to the APL.
- **Application Perspective Layer (APL)** first assesses the QoS request received from the UPL, and aims to satisfy the needs of the multimedia application by requesting the required services from the TPL.
- **Transmission Perspective Layer (TPL)** is responsible for negotiating with the network infrastructure to obtain appropriate communication services that can guarantee QoS.

Similarly, the various QoS protocols developed at the network infrastructure level need to be

Figure 2. Three layer QoS (TRAQS) model (Sharda, 1999)



able to communicate with the TPL to allow the user to specify the desired compromise between cost, quality, and temporal issues such as delay and jitter.

Quality, Cost, Temporal Triangle (QCTT)

In purchasing any goods or services one needs to find a compromise between three important factors: cost, quality, and time. While one would like to get the best quality at the least cost and in the shortest time, in practice, this is not possible. One must strike a compromise between these three factors. So far mobile communications systems have not come to grips with this reality. Future communication systems must provide users the ability to specify what quality and temporal aspects (such as delay and jitter) they want, and then systems should respond with the cost it would charge to provide that quality.

Only with a differentiated cost can the telecommunications service providers afford to deliver the required QoS. If the cost is too low, the network may be overwhelmed with traffic, and none of the users can then obtain the desired QoS. Over and above this, the network services provider may not be able to make profit. On the other hand, if the cost is too high, there will not be enough consumers using the service, once again making it difficult for the service provider to get return on investment. Whereas, by providing the user the ability to negotiate, the consumer and the service provider can both have a win-win situation; some

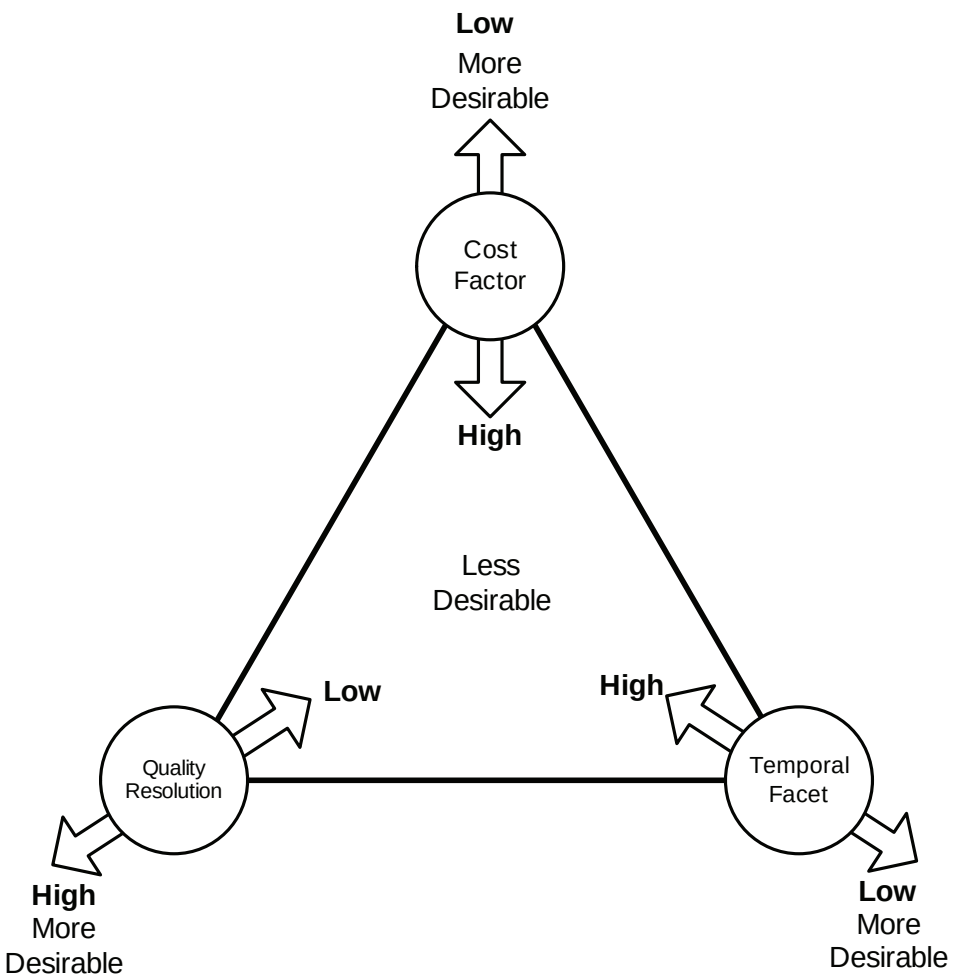
consumers pay high cost as they need higher quality, while other consumers can pay lower cost, as they have lower QoS requirements.

In the following sections, we first explain the concepts involved in the quality, cost, temporal triangle (QCTT), and then present an implementation of the same (Georgievski & Sharda, 2005a).

The three performance aspects—quality, cost, and time—are bound by a tri-partite dependency and thus can be modelled as a triangular relationship, as shown in Figure 3. The QCTT model embodies an inherent restriction on the delivery of QoS, that is, it is possible to achieve the more desirable parameter values only for two of the three performance aspects, while the third aspect must be forced to the less desirable value (Georgievski & Sharda, 2005b).

For example, if a user chooses to have high quality resolution (e.g., large image size, high frame rate), and, the more desirable, low temporal facet (e.g., low delay and jitter), then the cost factor has got to be high. By embedding the quality, cost temporal (QCTT) model in a user interface, we can provide the ability to dynamically man-

Figure 3. Quality, cost, temporal triangle (QCTT) model (Georgievski & Sharda, 2005a)



age QoS even while a multimedia session is in progress.

A multimedia communication session first needs to enter static QoS specifications, and then carry out dynamic QoS management as the session proceeds. An interface based on the QCTT model provides the ability to dynamically manage QoS. Such interfaces are described in the following sections.

Static QoS Specification

Figure 4 shows the user interface developed for negotiating static QoS prior to initiating a multimedia communication session. Using this interface, the user is able to specify the desired QoS, and then interactively negotiate with the system. It uses intuitive GUI elements such as a four colour system, a user status response, and a system status signalling system. These GUI elements allow the user to request the desired QoS, and get feedback if the network can deliver the same (Georgievski & Sharda, 2005a).

Dynamic QoS Management with QCTT

A dynamic QoS management interface is shown in Figure 5. This interface uses the QCTT model for re-negotiating QoS while a communication session is taking place. This is achieved by using three GUI elements: three sliders, buttons, and pivot point displacement. The system feedback GUI elements include: system QoS provision ring and values, and QCT threshold line (Georgievski & Sharda, 2005a).

To specify the desired QoS, the user moves the pivot point in the QCT triangle to a location which indicates the desired values for quality, cost, and temporal parameters.

The system provides visual feedback as follows:

1. **QoS Provision Ring** displays the current QoS parameter values that the system is able to provide.
2. **QoS Provision** values display the current numerical values set for QoS parameters.
3. **QCT Threshold Line** uses a three-colour scheme to provide feedback for displaying desirable and non-desirable values for each aspect.

This system has been tested and a usability analysis has been carried out on the same. While some improvements have been stipulated in the current implementation, overall, it received good assessment from the users (Georgievski & Sharda, 2005a).

Quality of Service on the Move

The ability to provide the requested QoS while roaming will be an important aspect in differentiating various mobile operators. This will determine their ability to hold on to their customers and therefore their revenue stream.

Providing QoS to a customer on the move is highly complex. Factors such as continuous handover, variable quality, dropout, and environmental factors make delivery of consistent QoS highly problematic. QoS provisioning has three main aspects: (1) resource reservation, (2) QoS routing protocol, and (3) Call admission control policy.

The integrated services (IntServ) framework developed under the RFC 1633 aims to provide customised QoS to individual applications (Aggélou, 2003). This is based on two aspects:

1. **Resource Reservation:** Each router needs to know the amount of buffer space and link bandwidth it needs to reserve for a session.
2. **Call Admission:** Each router determines the resources already committed to current sessions it is serving, before accepting the request from a new session.

Figure 4. Static QoS negotiation user interface (Georgievski & Sharda, 2005a)

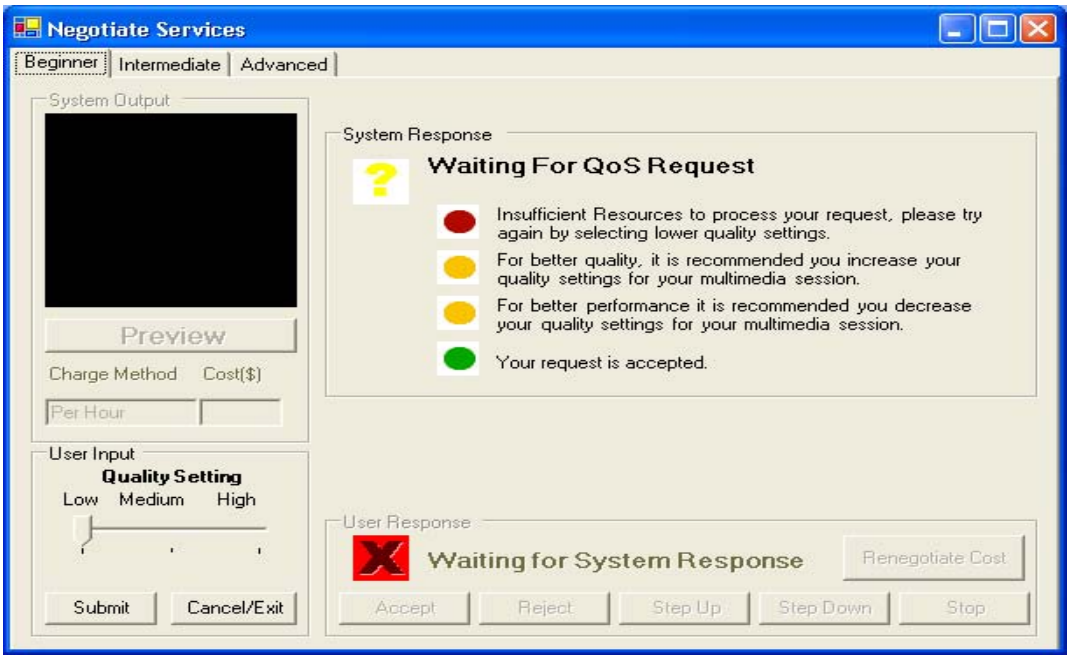
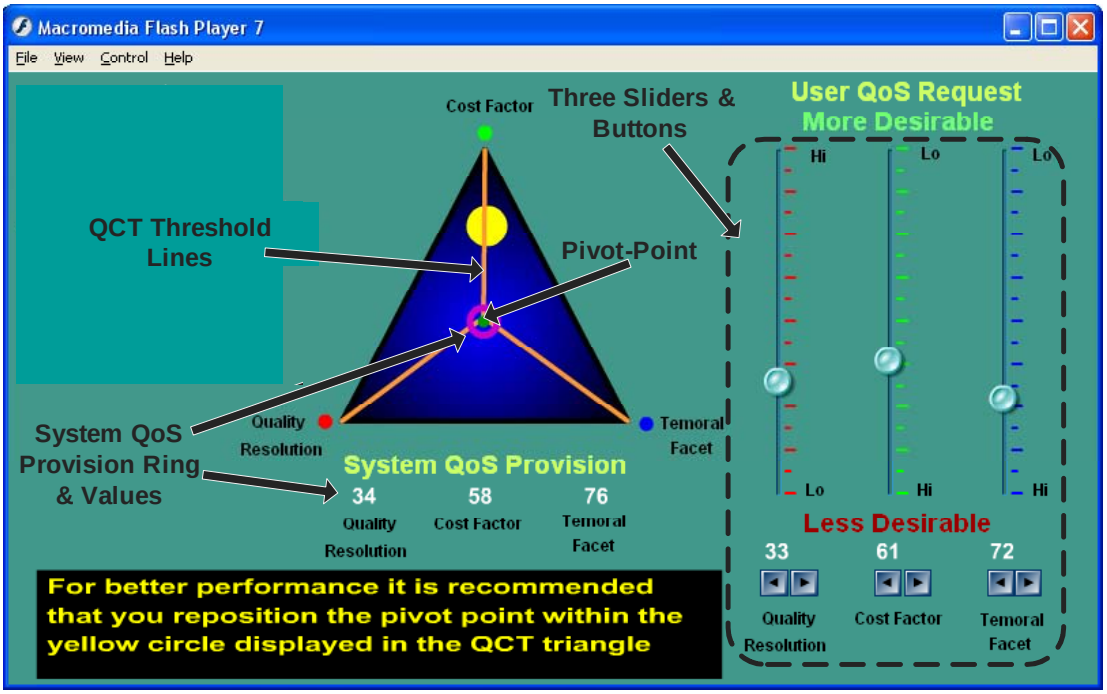


Figure 5. Dynamic QoS management user interface (Georgievski & Sharda, 2005a)



QoS routing is the most important protocol for mobile networks, the main objective specified for this protocol in the RFC 2386 are:

1. **Dynamic Determination of Feasible Paths:** This is based on policy and cost constraints.
2. **Optimisation of Resource Usage:** This requires state-dependent routing schemes.
3. **Graceful Performance Degradation:** This aspect compensates for transient inadequacies using the state-dependent routing scheme.

In summary, a mobile network needs the ability to reserve resources, ensure that a new call is admitted only if enough resources are available, choose the most suitable path to optimise the utilisation of resources, and provide graceful degradation in performance as resources become overloaded.

Quality of Services in Mobile Ad-Hoc Networks

Mobile ad-hoc networks are becoming an important area of investigation. As routing paths are not fixed in an ad-hoc network, QoS routing becomes an even more dynamic problem (Aggélou, 2004). In any ad-hoc network, a variety of routes with differing node capacity and power may be available to transmit data to the destination.

In general, not all routes are capable of providing the required QoS to satisfy the needs of the mobile users. Even when a route is selected that initially meets the user requirements, its error characteristics will not remain constant with time, due to the dynamic nature of routing and node placement in mobile ad-hoc networks. Therefore, ongoing re-routing will be required in an ad-hoc mobile network.

MOSQUITO: Mobile Quality of Service Provision in the Multi-Service Network

The MOSQUITO project, at the University College London, explored a microeconomic approach to resource allocation for providing QoS over multi-service network.

In this protocol, a base station sells bandwidth and QoS guarantees in small auctions to mobile terminals. A simple price setting / bidding function is used to determine the outcome of the auction. This research project aims to explore if:

- Microeconomics can be used for resource allocation.
- The performance of such a system can be measured.
- The algorithm creates a stable system or a chaotic one.
- Chaos can be characterised and controlled.

To use microeconomics for QoS provisioning, such questions need to be answered. Additionally, pricing functions need to be established using some simplifying assumptions; because, without simplifying heuristics, the juxtaposition of a myriad of factors such as pricing, routing, and quality selection will make real-time negotiations impossible.

FUTURE DIRECTIONS

There is no doubt that the future is heading towards mobile communications. And multimedia information will increasingly become the main traffic being transmitted, or blocked, on these networks. One solution to this problem is the so called “brute force” method: that is, “throwing” more bandwidth at the multimedia applications. However, experience shows that as more resources are made available without a viable economic model, the system ultimately gets overloaded.

Therefore, developing QoS systems that provide the user with the ability to negotiate with the network infrastructure are going to be of paramount importance.

Some of the developments in this area point to the following:

1. In the near future, mobile computing and communication systems will suffer from low bandwidth and low performance due to battery limitations.
2. Increasingly, mobile systems will provide higher bandwidth and combine different wireless technologies, such as high performance local wireless networks and wide area networks.
3. Third generation mobile systems will combine IP-traffic with traditional voice traffic.

The next generation of mobile networking technology is called 4G, or "3G and beyond" by IEEE (Aggélou & Tafazolli, 2001). In Japan, NTT DoCoMo is conducting tests under the 4G banner for 100 Mbps speeds with moving terminals, and 1 Gbps for stationary terminals. The first commercial release by NTT DoCoMo is expected in 2010. This technology aims to provide on demand high quality video and audio. 4G will use OFDM (orthogonal frequency division multiplexing), and also OFDMA (orthogonal frequency division multiple access) to better allocate network resources, and service multiple users simultaneously. Unlike the 3G networks, which use both circuit switching and packet switching, 4G will use packet switching only. Additionally, many QoS issues will be handled by developing new protocols. Nonetheless, the author contends that providing the ability to negotiate a compromise between cost, quality, and temporal aspects will remain an important issue.

CONCLUSION

Transmission of multimedia information over mobile networks is becoming increasingly important. New applications are in the offing if such multimedia information can be transmitted with the desired QoS. Text and still images do not pose much problem when transmitting these over mobile networks, as delay and delay variance do not adversely effect the operation of applications using text or still images. JPEG 2000 standard provides a marked improvement over the current standards such as JPEG, GIF, and PNG for still image transmission. Audio and video transmission, especially for full-duplex applications requiring real-time operation, poses the most demanding requirements for providing the desired QoS. While 3G networks, and 4G networks of the future, are capable of providing the required infrastructure for delivering multimedia content with the desired QoS, their user interfaces need to provide the ability to strike the desired balance between quality, cost, and temporal aspects.

ACKNOWLEDGMENT

The author would like to thank Dr. Mladen Georgievski for his useful suggestions and other contributions towards the preparation of this chapter.

REFERENCES

Aggélou, G. (2004). *Mobile ad hoc networks.* New York: McGraw-Hill Professional.

Aggélou, G., & Tafazolli, R. (2001). QoS support in 4th generation mobile multimedia ad hoc networks. *Proceedings of the Second International Conference on 3G Mobile Communication Technologies*, London, March 26-28 (pp. 412-416). London: Institute of Electrical Engineers.

Aggélou, G. N. (2003). An integrated platform for quality-of-service support in mobile multimedia clustered ad hoc networks. In M. Ilyas (Ed.), *The handbook of ad hoc wireless networks* (pp. 443-465). Boca Raton, FL: CRC Press, Inc.

Cheng, A., & Shang, F. (2005). Priority-driven coding of progressive JPEG images for transmission in real-time applications. *11th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'05),* Hong Kong, August 17-19 (pp. 129-134). Washington, DC: IEEE Computer Society.

Dietz, M., & Meltzer, S. (2002, July). CT-aacPlus: A state-of-the-art audio coding scheme. *EBU Technical Review,* (291), 1-7. Retrieved from http://www.ebu.ch/en/technical/trev/trev_291-dietz.pdf and http://www.ebu.ch/en/technical/trev/trev_index-digital.html

Dufaux, F., & Nicholson, D. (2004). JPWL: JPEG 2000 for wireless applications. Photonic devices and algorithms for Computing VI. In K. M. Iftekharuddin, & A. A. S. Awwal (Eds.), *Proceedings of the SPIE, 5558*, 309-318.

Georgievski, M., & Sharda, N. (2005a). Enhancing user experience for networked multimedia systems. *Proceedings of the 4th International Conference on Information Systems Technology and its Applications (ISTA2005*), Massey University, Palmerston North, New Zealand, May 23-25 (pp. 73-84). Bonn: Lecture Notes in Informatics (LNI), Gesellschaft für Informatik (GI).

Georgievski, M., & Sharda, N. (2005b). Implementation and usability of user interfaces for quality of service management. *Tencon'05: Proceedings of the Annual technical Conference of IEEE Region 10,* Australia, November 21-24. New Jersey: IEEE.

Liu, T., & Choudary, C. (2004). Content-aware streaming of lecture videos over wireless networks. *IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04),* Miami, FL, December 13-15 (pp. 458-465). Washington, DC: IEEE Computer Society.

Myers, D. (2004). *Mobile video telephony.* New York: McGraw-Hill Professional.

Navakitkanok, P., & Aramvith, S. (2004). Improved rate control for advanced video coding (AVC) standard under low delay constraint. *International Conference on Information Technology: Coding and Computing (ITCC'04), 2,* Las Vegas, NV, April 5-7 (p. 664). Washington, DC: IEEE Computer Society.

Rokou, F. P., & Rokos, Y. (2004). Integral laboratory for creating and delivery lessons on the Web based on a pedagogical content repurposing approach. *Fourth IEEE International Conference on Advanced Learning Technologies (ICALT'04),* Joensuu, Finland, August 30-September 1 (pp. 732-734). Washington, DC: IEEE Computer Society.

Santa-Cruz, D., Grosbois, R., & Ebrahimi, T. (2002). JPEG 2000 performance evaluation and assessment. *Signal Processing: Image Communication, 17*(1), 113-130.

Secker, A., & Taubman, D. S. (2004). Highly scalable video compression with scalable motion coding. *IEEE Transactions on Image Processing, 13*(8), 1029-1041.

Sharda, N. (1999). *Multimedia information networking.* New Jersey: Prentice Hall.

Sharda, N., & Georgievski, M. (2002). A holistic quality of service model for multimedia communications. *International Conference on Internet and Multimedia Systems and Applications (IMSA2002),* Kaua'i, Hawaii, August 12-14 (pp. 282-287). Calgary, Alberta, Canada: ACTA Press.

Smith, J. R., & Jabri, M. A. (2004). The 3G-324M protocol for conversational video telephony. *IEEE MultiMedia, 11*(3), 102-105.

Tabesh, A., Bilgin, A., Krishnan, K., & Marcellin, M. W. (2005). JPEG2000 and motion JPEG2000 content analysis using codestream length information. *Proceedings of The Data Compression Conference (DCC'05),* Snowbird, UT, March 29-31 (pp. 329-337). Washington, DC: IEEE Computer Society.

Taubman, D., & Marcellin, M. (2002). *JPEG2000: Image compression fundamentals, standards and practice*. Netherlands: Kluwer Academic Publishers.

ENDNOTE

[1] CIF: Common Intermediate Format. A video format used in videoconferencing systems. It is part of the ITU H.261 videoconferencing standard, and specifies a data rate of 30 frames per second (fps), with each frame containing 288 lines and 352 pixels per line. Other CIF based standards include: QCIF - Quarter CIF (176x144), SQCIF - Sub quarter CIF (128x96), 4CIF - 4 x CIF (704x576), and 16CIF - 16 x CIF (1408x1152).

This work was previously published in Mobile Multimedia Communications: Concepts, Applications, and Challenges, edited by G. Karmakar & L. Dooley, pp. 45-63, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.5
Perceptual Semantics

Andrea Cavallaro
Queen Mary University of London, UK

Stefan Winkler
National University of Singapore and Genista Corporation, Singapore

ABSTRACT

The design of image and video compression or transmission systems is driven by the need for reducing the bandwidth and storage requirements of the content while maintaining its visual quality. Therefore, the objective is to define codecs that maximize perceived quality as well as automated metrics that reliably measure perceived quality. One of the common shortcomings of traditional video coders and quality metrics is the fact that they treat the entire scene uniformly, assuming that people look at every pixel of the image or video. In reality, we focus only on particular areas of the scene. In this chapter, we prioritize the visual data accordingly in order to improve the compression performance of video coders and the prediction performance of perceptual quality metrics. The proposed encoder and quality metric incorporate visual attention and use a semantic segmentation stage, which takes into account certain aspects of the cognitive behavior of people when watching a video. This semantic model corresponds to a specific human abstraction, which need not necessarily be characterized by perceptual uniformity. In particular, we concentrate on segmenting moving objects and faces, and we evaluate the perceptual impact on video coding and on quality evaluation.

INTRODUCTION

The development of new compression or transmission systems is driven by the need of reducing the bandwidth and storage requirements of images and video while increasing their perceived visual quality. Traditional compression schemes aim at minimizing the coding residual in terms of mean squared error (MSE) or peak signal-to-noise ratio (PSNR). This is optimal from a purely mathematical but not a perceptual point of view. Ultimately, perception is the more appropriate and more relevant benchmark. Therefore, the objective must be to define a codec that maximizes perceived visual quality such that it produces better quality at the same bit rate as a traditional encoder or the same visual quality at a lower bit rate (Cavallaro, 2005b).

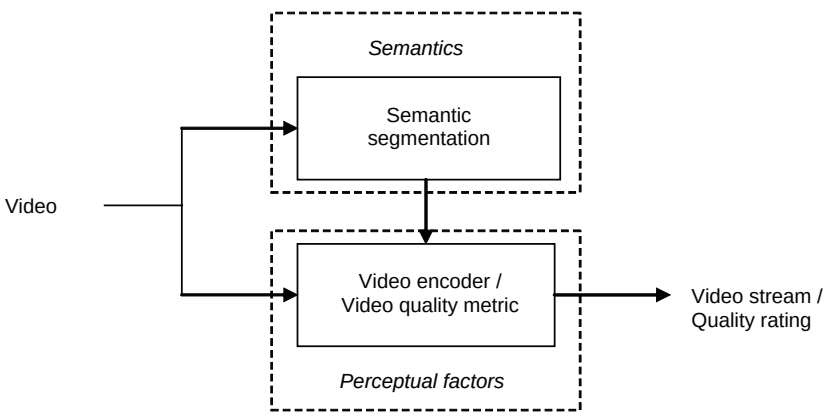
In addition to achieving maximum perceived quality in the encoding process, an important concern for content providers is to guarantee a certain level of quality of service during content distribution and transmission. This requires reliable methods of quality assessment. Although subjective viewing experiments are a widely accepted method for obtaining meaningful quality ratings for a given set of test material, they are necessarily limited in scope and do not lend themselves to monitoring and control applications, where a large amount of content has to be evaluated in real-time or at least very quickly. Automatic quality metrics are desirable tools to facilitate this task. The objective here is to design metrics that predict perceived quality better than PSNR (Winkler, 2005a).

One of the common shortcomings of traditional video coders and quality metrics is the fact that they treat the entire scene uniformly, assuming that people look at every pixel of the image or video. In reality, we focus only on particular areas of the scene, which has important implications on the way the video should be analyzed and processed.

In this chapter, we take the above observations into account and attempt to emulate the human visual system. The idea is to prioritize the visual data in order to improve the compression performance of video coders and the prediction performance of perceptual quality metrics. The proposed encoder and quality metric incorporate visual attention and use a semantic segmentation stage (Figure 1). The semantic segmentation stage takes into account some aspects of the cognitive behavior of people when watching a video. To represent the semantic model of a specific cognitive task, we decompose each frame of the reference sequence into sets of mutually-exclusive and jointly-exhaustive segments. This semantic model corresponds to a specific human abstraction, which need not necessarily be characterized by perceptual uniformity. Since the semantics (i.e., the meaning) are defined through human abstraction, the definition of the semantic partition depends on the task to be performed. In particular, we will concentrate on segmenting moving objects and faces, and we will evaluate the perceptual impact on video coding and on quality evaluation.

The chapter is organized as follows: The section "Cognitive Behavior" discusses the factors influencing the cognitive behavior of people watching a video. The section "Semantic Segmentation" introduces the segmentation stage that generates a semantic partition to be used in video coding and quality evaluation. In "Perceptual Semantics for Video Coding" and "Perceptual Semantics for Video Quality Assessment", we describe how

Figure 1. Flow diagram of the encoder and the quality metric that incorporate factors influencing visual attention and use a semantic segmentation stage



the cognitive behavior can be incorporated into a video coder and a quality metric, respectively. Moreover, the compression performance of the proposed encoder and the prediction performance of the proposed metrics are discussed. In the final section, we draw some conclusions and describe the directions of our current work.

COGNITIVE BEHAVIOR

Visual Attention

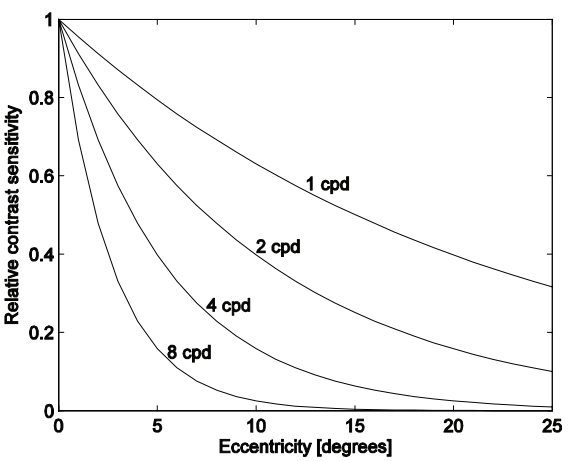
When watching images or video, we focus on particular areas of the scene. We do not scan a scene in raster fashion; instead, our visual attention tends to jump from one point to another. These so-called saccades are driven by a fixation mechanism, which directs our eyes towards objects of interest (Wandell, 1995). Saccades are high-speed eye movements that occur at a rate of 2-3 Hz. We are unaware of these movements because the visual image is suppressed during saccades.

Studies have shown that the direction of gaze is not completely idiosyncratic to individual viewers. Instead, a significant number of viewers will focus on the same regions of a scene (Endo, 1994; Stelmach, 1994). Yarbus (1967) demonstrated in his classic experiments that the saccadic patterns depend on the visual scene as well as the cognitive task to be performed. In other words,"we do not see, we look" (Bajcsy, 1988, p. 1). We focus our visual attention according to task at hand and the scene content.

The reason why this behavior should be taken into account in video encoding and quality assessment is that our visual acuity is not uniform across the entire visual field. In general, visual acuity is highest only in a relatively small cone around the optical axis (the direction of gaze) and decreases with distance from the center. This is partly due to the deterioration of the optical quality of the eye towards the periphery and partly due to the layout of the retina (Banks, Sekuler, & Anderson, 1991). The central region of the retina around the optical axis is called the fovea. It contains the highest density of cones, which are the photoreceptors responsible for vision under usual image or video viewing conditions. Outside of the fovea, the cone density decreases rapidly. This explains why vision is sharp only around the present focus of attention, and the perception of the peripheral field of vision is blurred. In other words, contrast sensitivity is reduced with increasing eccentricity (Robson & Graham, 1981).[1] This effect is also frequency-dependent, as shown in Figure 2.

While the processes governing perception are not completely understood, many different factors contributing to visual attention have been identified (Wolfe, 1998). These include simple stimulus properties such as contrast (regions with high contrast attract attention), size (large objects are more interesting than small ones), orientation, shape (edge-like features or objects with corners are preferred over smooth shapes), hue and intensity (an object with very different color from the background or specific bright colors stands out), or

Figure 2. Loss of spatial contrast sensitivity as a function of eccentricity and stimulus frequency (measured in cycles per degree [cpd]), based on a model by Geisler and Perry (1998)



flicker and motion (see below). In general, an object or feature that stands out from its surroundings in terms of any of the earlier-mentioned factors is more likely to attract our attention. Additionally, the location (most viewers focus on the center of a scene), foreground/background relationship, and context (the task, personal interest, or motivation) influence visual attention. Finally, moving objects and people, in particular their faces (eyes, mouth) and hands, draw our attention.

Computational Models of Attention

While most vision models and quality metrics are limited to lower-level aspects of vision, the cognitive behavior of people when watching video cannot be ignored. Cognitive behavior may differ greatly between individuals and situations, which makes it very difficult to generalize. Furthermore, little is known about the relative importance given to the factors listed above by the human visual system. Nonetheless, based on the above studies and factors contributing to visual attention, a variety of computational models of attention have been developed (Itti & Koch, 2001; Wolfe, 1998). Maeder, Diederich, and Niebur (1996) proposed constructing an importance map for a sequence as a prediction for the focus of attention, taking into account perceptual factors such as edge strength, texture energy, contrast, color variation, and homogeneity. Osberger and Rohaly (2001) developed a segmentation-based model with motion estimation and skin detection, taking into account color, contrast, size, shape, location, and foreground. It was calibrated and tested with eye movement data gathered from an experiment using a large database of still and video scenes. Recently, Navalpakkam and Itti (2005) demonstrated a model that uses task-specific keywords to find relevant objects in a scene and extracts their low-level features to build a visual map of task-relevance.

In this chapter, we focus on two important high-level aspects, namely faces and moving objects.

Attraction of Faces and Moving Objects

People and especially faces are among the objects attracting the most attention.. If there are faces of people in a scene, we will look at them immediately. Furthermore, because of our familiarity with people's faces, we are very sensitive to distortions or artifacts occurring in them. The importance of faces is underlined by a study of image appeal in consumer photography (Savakis, 2000). People in the picture and their facial expressions are among the most important criteria for image selection.

In a similar manner, viewers may also track specific moving objects in a scene. In fact, motion (in particular, in the periphery) tends to attract the viewers' attention. The spatial acuity of the human visual system depends on the velocity of the image on the retina: As the retinal image velocity increases, spatial acuity decreases. The visual system addresses this problem by tracking moving objects with smooth-pursuit eye movements, which minimizes retinal image velocity and keeps the object of interest on the fovea. Smooth pursuit works well even for high velocities, but it is impeded by large accelerations and unpredictable motion (Eckert, 1993). On the other hand, tracking a particular movement will reduce the spatial acuity for the background and objects moving in different directions or at different velocities. An appropriate adjustment of the spatio-temporal contrast sensitivity function (CSF) as outlined by Daly (1998) to account for some of these sensitivity changes can be considered as a first step in modeling such phenomena.

Based on the observations of this section, the proposed video coder and perceptual quality metric take into account both low-level and high-level aspects of vision. To achieve this, a segmentation stage is added to the video coder and to the quality metric to find regions of interest. The segmentation output then guides a pre-processing step in coding and a pooling process in video quality evaluation by giving more weight to the regions with semantically higher importance.

Figure 3. (a) Example of face detection result on the test sequence Susie and (b) corresponding segmentation mask (white: foreground, black: background)

(a) (b)

SEMANTIC SEGMENTATION

The high-level contribution to the cognitive behavior of people when watching a video is taken into account by means of semantic segmentation. To represent the semantic model of a specific cognitive task, we decompose each frame of the sequence into sets of mutually-exclusive and jointly-exhaustive segments. In general, the topology of this semantic partition cannot be expressed using homogeneity criteria, because the elements of such a partition do not necessarily possess invariant properties. As a consequence, some knowledge of the objects we want to segment is required. We will consider two cases of such *a priori* information, namely segmentation of faces and segmentation of moving objects. The final partition will be composed of foreground areas and background areas of each image.

Color segmentation and feature classification can be exploited to segment faces of people. A number of relatively robust algorithms for face segmentation are based on the fact that human skin colors are confined to a narrow region in the chrominance (C_b, C_r) plane (Gu, 1999), when the global illumination of the scene does not change significantly. Otherwise, methods based on tracking the evolution of the skin-color distribution at each frame based on translation, scaling, and rotation of skin color patches in the color space can be used (Sigal, 2004). One of the limits of approaches based on color segmentation is that the resulting partition may include faces as well as body parts. To overcome this problem, a combination of color segmentation with facial feature extraction (Hsu, 2002) can be used. Other approaches use feature classifiers only. Viola and Jones (2004) proposed a face detector based on a cascade of simple classifiers and on the integral image. This detector is a multi-stage classification that works as follows. First, features similar to Haar basis functions are extracted from the gray-level integral image. Next, a learning step called AdaBoost is used to select a small number of relevant features. This pruning process selects weak classifiers that depend on one feature only. Finally, the resulting classifiers are combined in a cascade structure. With such an approach, a face cannot be reliably detected when it appears small or not frontal. However, in such a situation the face in general does not attract viewers' attention as much as a frontal or large face. Therefore, this limitation will not affect the proposed approach in a significant way. Figure 3 shows an example of face detection for the test sequence Susie.

To segment moving objects, motion information is used as semantics. The motion of an object is usually different from the motion of background and other surrounding objects. For this reason, many extraction methods make use of motion information in video sequences to segment objects (Cavallaro, 2004a). Change detection is a typical

tool used to tackle the problem of object segmentation based on motion. Different change detection techniques can be employed for moving camera and static camera conditions. If the camera moves, change detection aims at recognizing coherent and incoherent moving areas. The former correspond to background areas, the latter to video objects. If the camera is static, the goal of change detection is to recognize moving objects (foreground) and the static background. The semantic segmentation discussed here addresses the static camera problem and is applicable in the case of a moving camera after global motion compensation.

The change detector decides whether in each pixel position the foreground signal corresponding to an object is present. This decision is taken by thresholding the frame difference between the current frame and a frame representing the background. The frame representing the background is dynamically generated based on temporal information (Cavallaro, 2001). The thresholding aims at discarding the effect of the camera noise after frame differencing. A locally adaptive threshold, $\tau(i,j)$, is used that models the noise statistics and applies a significance test. To this end, we want to determine the probability that frame difference at a given position *(i,j)* is due to noise, and not to other causes. Let us suppose that there is no moving object in the frame difference. We refer to this hypothesis as the null hypothesis, H_0. Let *g(i,j)* be the sum of the absolute values of the frame difference in an observation window of *q* pixels around *(i,j)*. Moreover, let us assume that the camera noise is additive and follows a Gaussian distribution with variance σ. Given H_0, the conditional probability density function (pdf) of the frame difference follows a χ^2_q distribution with *q* degrees of freedom defined by:

$$f(g(i,j) \mid H_o) = \frac{1}{2^{q/2}\sigma^q \Gamma(q/2)} g(i,j)^{(q-2)/2} e^{-g(i,j)^2/2\sigma^2}$$

$\Gamma(\cdot)$ is the Gamma function, which is defined as $\Gamma(x+1) = x\Gamma(x)$, and $\Gamma(1/2) = \sqrt{\pi}$. To obtain a good trade-off between robustness to noise and accuracy in the detection, we choose *q=25* (5-by-5 window centered in *(i,j)*). It is now possible to derive the significance test as:

$$P\{g(i,j) \geq \tau(i,j) \mid H_0\} = \frac{\Gamma(q/2, g(i,j)^2/2\sigma^2)}{\Gamma(q/2)} .$$

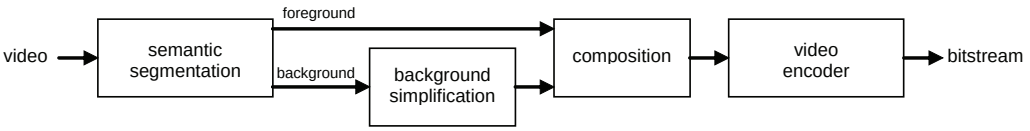
When this probability is smaller than a certain significance level, α, we consider that H_0 is not satisfied at the pixel position *(i,j)*. Therefore we label that pixel as belonging to a moving object. The significance level α is a stable parameter that does not need manual tuning over a sequence or for different sequences. Experimental results indicate that valid values fall in the range from 10^{-2} to 10^{-6}.

Figure 4. Example of moving object segmentation. (a) Original images; (b) moving object segmentation mask (white: foreground, black: background)

(a)

(b)

Figure 5. Using semantic segmentation for perceptual video coding



An example of a moving object segmentation result is shown in Figure 4.

The video partitions generated with semantic segmentation are then used in the video encoding process and in the quality evaluation metric as described in the following sections.

PERCEPTUAL SEMANTICS FOR VIDEO CODING

A video encoder that exploits semantic segmentation facilitates the improvement of video compression efficiency in terms of bandwidth requirements as well as visual quality at low bit rates. Using semantic decomposition, the encoder may adapt its behavior to code relevant and non-relevant portions of a frame differently. This adaptation can be achieved with a traditional frame-based encoder by semantically pre-filtering the video prior to coding (Figure 5).

Video Coders

The semantic pre-filtering is obtained by exploiting semantics in a traditional frame-based encoding framework, such as MPEG-1. The use of the decomposition of the scene into meaningful objects prior to encoding, referred here as semantic pre-filtering, helps support low-bandwidth transmission. The areas belonging to the foreground class are used as the region of interest. The areas not included in the region of interest are lowered in importance by using a low-pass filter. The latter solution simplifies the information in the background, while still retaining essential contextual information. The simplification of the background allows for a reduction of information to be coded. The use of a simplified background so as to enhance the relevant objects aims at taking advantage of the task-oriented behavior of the human visual system for improving compression ratios. The work reported by Bradley (2003) demonstrates that an overall increase in image quality can be obtained when the increase

Figure 6. Simplification of contextual information: (a) Sample frame from the test sequence Hall Monitor; (b) simplification of the whole frame using a low-pass filter; (c) selective low-pass filtering based on semantic segmentation results allows one to simplify the information in the background while still retaining essential contextual information

(a) (b) (c)

in quality of the relevant areas of an image more than compensates for the decrease in quality of the image background. An example of this solution, which aims at mimicking the blur occurring in the retina due to foveation (Itti, 2004) using high-level semantic cues, is reported in Figure 6.

Another way to take into account less relevant portions of an image before coding is to take advantage of the specifics of the coding algorithm. In the case of block-based coding, each background macro-block can be replaced by its DC value only. While this approach also has the effect of frequency reduction and loss of detail, it may lead to unacceptably strong blocking artifacts in the video.

An alternative approach is using object-based encoding. In such a case, the encoder needs to support the coding of individual video objects, such as for MPEG-4, object-based mode (Sikora, 1997). With this solution, each video object is assigned to a distinct object class, according to its importance in the scene. The encoding quality can be set depending on the object class: the higher the relevance, the higher the encoding quality. One advantage of this approach is the possibility of controlling the sequencing of objects. Video objects may be encoded with different degrees of compression, thus allowing better granularity for the areas in the video that are of more interest to the viewer. Moreover, objects may be decoded in their order of priority, and the relevant content can be viewed without having to reconstruct the entire image. Finally, sprite coding could be used when an image representing the background is sent to the receiver once and then objects are encoded and composed with an appropriate portion of the background at the receiver side. However, these solutions require that objects are tracked after

Figure 7. Comparison between standard MPEG–1 encoding and perceptually pre-filtered MPEG–1 encoding at 150 kbit/s. Frame 190 of Hall Monitor: (a) MPEG–1 encoded; (b) MPEG–1 encoded after perceptual pre-filtering; Frame 44 of Highway (c) MPEG–1 encoded; (d) MPEG–1 encoded after perceptual pre-filtering. It is possible to notice that coding artifacts are less disturbing on objects of interest in (b) and (d) than in (a) and (c), respectively.

(a) (b)

(c) (d)

segmentation (Cavallaro, 2005a), thus increasing the complexity of the approach.

Results

The results presented in this section illustrate the impact of semantic pre-filtering on the encoding performance of a frame-based coder. Sample results are shown from the MPEG-4 test sequence Hall Monitor and from the MPEG-7 test sequence Highway. Both sequences are in CIF format at 25 Hz. The background is simplified using a Gaussian 9x9 low-pass filter with μ=0 and σ=2, where μ and σ are the mean and standard deviation of the filter, respectively. The TMPGEnc 2.521.58.169 encoder using constant bit-rate (CBR) rate control was used for the encoding.

Figure 7 shows a sample frame from each test sequence coded with MPEG-1 at 150 kbit/s with and without semantic pre-filtering. Figure 8 shows magnified excerpts of both test sequences coded with MPEG–1 at 150 kbit/s. Figure 8(a, b) shows a blue truck entering the scene at the beginning of the Highway sequence. Coding artifacts are less disturbing on the object in Figure 8(b) than in Figure 8(a). Moreover, the front-left wheel of the truck is only visible with semantic pre-filtering, Figure 8(b). Similar observations can be made for Figure 8(c, d), which shows the person that carries a monitor in the Hall Monitor sequence. The amount of coding artifacts is notably reduced by semantic pre-filtering as shown in Figure 8(d). In particular, the person's facial features and clothes are clearly visible in Figure 8(d), whereas they are corrupted by coding artifacts in Figure 8(c).

PERCEPTUAL SEMANTICS FOR VIDEO QUALITY ASSESSMENT

A quality measure based on semantic segmentation is useful for end-to-end communication architectures aiming at including perceptual requirements for an improved delivery of multimedia information. Predicting subjective ratings using an automatic visual quality metric with higher accuracy than peak signal-to-noise ratio (PSNR) has been the topic of much research in recent years. However, none of today's metrics quite achieves the reliability of subjective experiments.

Two approaches for perceptual quality metric design can be distinguished (Winkler, 2005b): One

Figure 8. Details of sequences encoded at 150 kbit/s: (a) Frame 44 of Highway MPEG–1 encoded; (b) Frame 44 of Highway MPEG–1 encoded after perceptual pre-filtering (c) Frame 280 of Hall Monitor MPEG–1 encoded; (d) Frame 280 of Hall Monitor MPEG–1 encoded after perceptual pre-filtering; notice that the front left wheel of the truck is visible with semantic pre-filtering only (b). The person's facial features and clothes are clearly visible in (d), whereas they are corrupted by coding artifacts in (c).

(a) (b) (c) (d)

class of metrics implements a general model of low-level visual processing in the retina and the early visual cortex. Metrics in this class typically require access to the reference video for difference analysis. The other class of metrics looks for specific features in the image, for example, compression artifacts arising from a certain type of codec, and estimates their annoyance.

To demonstrate the use of perceptual semantics with both classes of metrics, we describe a specific implementation from each class here. The first is a full-reference perceptual distortion metric (PDM) based on a vision model, and the second is a no-reference video quality metric based on the analysis of common artifacts. These are then combined with perceptual semantics to achieve a better prediction performance (Cavallaro & Winkler, 2004).

Full-Reference Quality Metric

The full-reference PDM is based on a contrast gain control model of the human visual system that incorporates spatial and temporal aspects of vision as well as color perception (Winkler, 1999). A block diagram of this metric is shown in Figure 9. The metric requires both the test sequence and the corresponding reference sequence as inputs.

The input videos are first converted from YUV or RGB to an opponent-color space, which is closer to the perceptual representation of color information. Each of the resulting three components is then subjected to a spatio-temporal decomposition, which is implemented as a filter bank. This emulates the different mechanisms in the visual system, which separate the input signals according to their color, frequency, orientation, and other characteristics. The sensitivity to the information in every channel is different, so the channels are weighted according to contrast sensitivity data.

The subsequent contrast gain control stage models pattern masking, which is one of the most critical components of video quality assessment, because the visibility of distortions is highly dependent on the local background. Masking is strongest between stimuli located in the same perceptual channel, and becomes weaker for stimuli with differing characteristics. The perceptual decomposition allows us to take many of these intra- and inter-channel masking effects into account. Within the process of contrast gain control, masking occurs through the inhibitory effect of the various perceptual channels.

At the output of the contrast gain control stage, the differences between the reference and the test video are computed for each channel and then combined ("pooled") into a distortion measure or quality rating according to the rules of probability summation.

No-Reference Quality Metric

The no-reference quality metric estimates visual quality based on the analysis of common coding and transmission artifacts found in the video (Win-

Figure 9. Block diagram of the perceptual distortion metric (PDM)

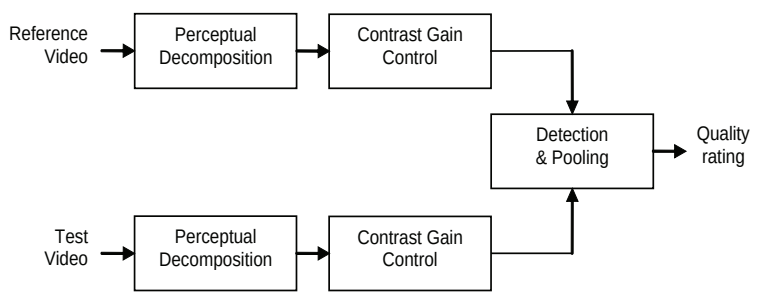
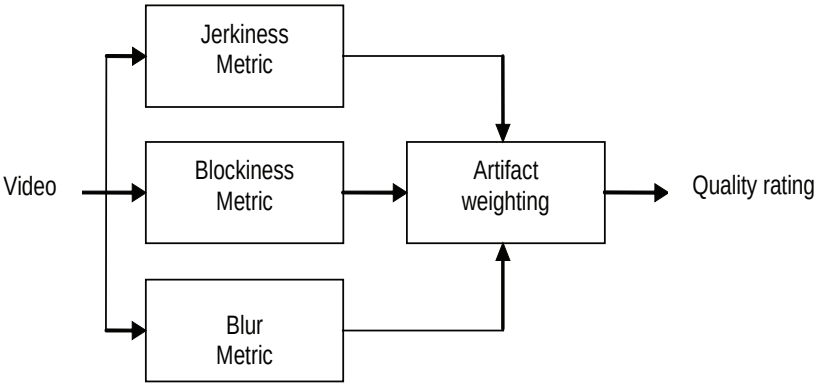


Figure 10. Block diagram of no-reference quality metric



kler & Campos, 2003). These artifacts include blockiness, blur, and jerkiness. A block diagram of this metric is shown in Figure 10. It does not need any information about the reference sequence.[2] The use of a no-reference metric is particularly interesting here because semantic segmentation does not require a reference video either.

The blockiness metric looks for typical block patterns in a frame. These block patterns are an artifact common to DCT block-based image and video compression methods such as JPEG or MPEG. Blocks often form regular horizontal or vertical structures in an image, which appear as characteristic peaks in the spectrum of spatial difference signals. If the blocks are not regular, each block has to be identified individually based on its texture and boundaries. The spectral power peaks and the boundary contrasts are used to compute a measure of overall blockiness.

Blur is a perceptual measure of the loss of fine detail in the image. It is due to the attenuation of high frequencies at some stage of the recording or encoding process. The blur metric is based on the assumption that edges, which often represent object contours, are generally sharp. Compression has a smearing effect on these edges. The blur metric thus looks for significant edges in an image and measures their width. This local edge smear is then averaged over the entire image for a global estimate of blur.

The jerkiness metric is a temporal metric dedicated to measuring the motion rendition in the video. Jerkiness is a perceptual measure of frozen pictures or motion that does not look smooth. The primary causes of jerkiness are network congestion and/or packet loss. It can also be introduced by the encoder dropping or repeating frames in an effort to achieve the given bit-rate constraints. The jerkiness metric takes into account the video frame rate as well as the amount of motion activity in the video.

Both of these metrics can make local quality measurements in small sub-regions over a few frames in every video. The process of combining these low-level contributions into an overall quality rating is guided by the result of the semantic segmentation stage as shown in Figure 11. The metrics described here attempt to emulate the human visual system to prioritize the visual data in order to improve the prediction performance of a perceptual distortion metric. The resulting semantic video quality metric thus incorporates both low-level and high-level aspects of vision. Low-level aspects are inherent to the metric and include color perception, contrast sensitivity, masking and artifact visibility. High-level aspects take into account the cognitive behavior of an observer when watching a video through semantic segmentation.

Figure 11. Quality assessment using perceptual semantics

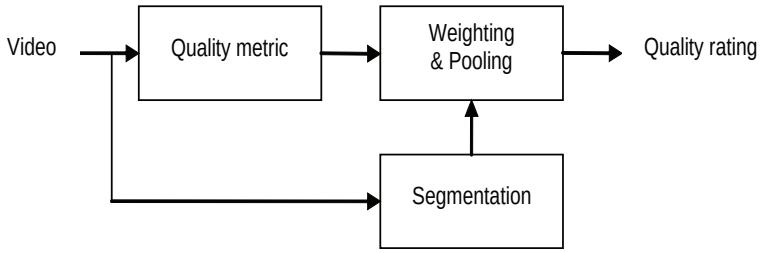
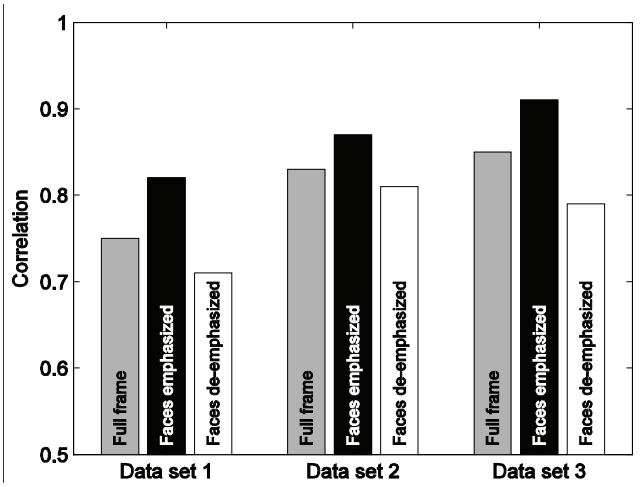


Figure 12. Sample frames from selected test sequences

Figure 13. Prediction performance with and without segmentation; correlations are shown for the metrics applied uniformly across the full frame (gray bars), with an emphasis on the areas resulting from face segmentation (black bars), and the complementary emphasis (white bars)



Evaluation and Results

To evaluate the improvement of the prediction performance due to face segmentation, we compare the ratings of the regular full-frame metrics with those of the segmentation-supported metrics for different data sets. We focus our evaluation on the special case when faces attract the attention of an observer. We quantify the influence of variations of quality of the relevant parts of the content on perceived video quality.

We used test material and subjective ratings from three different subjective testing databases:

1. VQEG Phase I database (VQEG, 2000): This database comprises mainly TV material with 16 test conditions. Three relevant scenes were selected from this database to evaluate the full-reference PDM.
2. PC video database (Winkler, 2001): This database was created with CIF-size video and various DirectShow codecs at bit-rates of 1-2 Mb/s, for a total of eight test conditions. We picked two scenes from this database to evaluate the full-reference PDM.
3. Internet streaming database (Winkler & Campos, 2003): This database contains clips encoded with MPEG-4, Real Media and Windows Media at 256 and 512 kb/s, as well as some packet loss (seven conditions in total). Four scenes from this database were used. Due to the test conditions here, these sequences cannot be properly aligned with the reference. Therefore, we use this set for the evaluation of our no-reference metric.

The scenes we selected from these databases contain faces at various scales and with various amounts of head and camera movements. Some examples are shown in Figure 12.

The results of the evaluation for our three data sets are shown in Figure 13. Segmentation generally leads to a better agreement between the metric's predictions and the subjective ratings. The trend is the same for all three data sets, which indicates that face segmentation is useful for augmenting the predictions of quality metrics. Moreover, giving lower weights to faces leads to a reduced prediction performance. This observation also supports the conclusion that semantic segmentation helps predicting perceived quality. As expected, the improvement is most noticeable for the scenes where faces cover a substantial part of the frame. Segmentation is least beneficial for sequences in which the faces are quite small and the distortions in the background introduced by some test conditions are more annoying to viewers than in other regions (as is the case with data set 2).

SUMMARY

We analyzed factors influencing visual attention and how they can be incorporated in a video encoder and in a perceptual distortion metric using segmentation. We looked at the special cases of moving object and face segmentation, and we evaluated the performance improvement of a frame-based encoder and video quality metrics combined with a segmentation stage.

In video coding, the advantages of segmentation support have been demonstrated at low-bit rates with test sequences containing moving objects. In quality evaluation, the advantages of segmentation support have been demonstrated with test sequences showing human faces, resulting in better agreement of the predictions of a perceptual quality metric with subjective ratings. Even if the segmentation support is adding complexity to the overall system, the algorithms used for the detection of faces and moving objects are running in real-time on standard personal computers. Moreover, the proposed perceptual video coding method does not add any overhead to the bitstream, thus allowing interoperability with existing decoders.

Current work includes the extension of the technique and results to the evaluation of multimedia information, including graphics, audio, and video for applications such as augmented reality and immersive spaces.

REFERENCES

Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE, 76*(8), 996-1005.

Banks, M. S., Sekuler, A. B., & Anderson, S. J. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America, A 8*, 1775-1787.

Bradley, A. P., & Stentiford, F. W. M. (2003). Visual attention for region of interest coding in JPEG-2000. *Journal of Visual Communication and Image Representation, 14*, 232-250.

Cavallaro, A., & Ebrahimi, T. (2001). Video object extraction based on adaptive background and statistical change detection. *Proceedings of SPIE Visual Communications and Image Processing, Vol. 4310* (pp. 465-475). San Jose, CA: SPIE.

Cavallaro, A., & Ebrahimi, T. (2004). Interaction between high-level and low-level image analysis for semantic video object extraction. *Journal on Applied Signal Processing, 2004*(6), 786-797.

Cavallaro, A., Steiger, O., & Ebrahimi, T. (2005). Tracking video objects in cluttered background. *IEEE Trans. Circuits and Systems for Video Technology, 15*(4), 575-584.

Cavallaro, A., & Steiger, O., & Ebrahimi, T. (in press). Semantic video analysis for adaptive content delivery and automatic description. *IEEE Trans. Circuits and Systems for Video Technology, 15*(19), 1200-1209.

Cavallaro, A., & Winkler, S. (2004). Segmentation-driven perceptual quality metrics. *Proceedings of the International Conference on Image Processing* (pp. 3543-3546). Singapore: IEEE.

Daly, S. (1998). Engineering observations from spatio-velocity and spatio-temporal visual models. *Proceedings of SPIE Human Vision and Electronic Imaging*: *Vol. 3299* (pp. 180-191). San Jose, CA: SPIE.

Eckert, M. P., & Buchsbaum, G. (1993). The significance of eye movements and image acceleration for coding television image sequences. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 89-98). Cambridge, MA: MIT Press.

Endo, C., Asada, T., Haneishi, H., & Miyake, Y. (1994). Analysis of the eye movements and its applications to image evaluation. *Proceedings of the Color Imaging Conference* (pp. 153-155). Scottsdale, AZ: IS&T/SID.

Geisler, W. S., & Perry, J. S. (1998). A real-time foveated multi-resolution system for low-bandwidth video communication. *Proceedings of SPIE Human Vision and Electronic Imaging*: *Vol. 3299* (pp. 294-305).

Gu, L., & Bone, D. (1999). Skin color region detection in MPEG video sequences. *Proceedings of the International Conference on Image Analysis and Processing* (pp. 898-903). Venice, Italy: IAPR.

Hsu, L., Abdel-Mottaleb, M., & Jain, A. K. (2002). Face detection in color images. *IEEE Trans. Pattern Analysis and Machine Intelligence, 24*(5), 696-706.

Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans. Image Processing, 13*(10), 1304-1318.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

Maeder, A., Diederich, J., & Niebur, E. (1996). Limiting human perception for image sequences. *Proceedings of SPIE Human Vision and Electronic Imaging*: *Vol. 2657* (pp. 330-337). San Jose, CA: SPIE.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*, 205-231.

Osberger, W., & Rohaly, A. M. (2001). Automatic detection of regions of interest in complex video sequences. *Proceedings of SPIE Human Vision*

and Electronic Imaging: Vol. 4299 (pp. 361-372). San Jose, CA: SPIE.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21,* 409-418.

Savakis, A. E., Etz, S. P., & Loui, A. C. (2000). Evaluation of image appeal in consumer photography. *Proceedings of SPIE Human Vision and Electronic Imaging: Vol. 3959* (pp. 111-120). San Jose, CA: SPIE.

Sigal, L., Sclaroff, S., & Athitsos, V. (2004). Skin color-based video segmentation under time-varying illumination. *IEEE Trans. Pattern Analysis and Machine Intelligence, 26*(7), 862-877.

Sikora, T. (1997). The MPEG-4 video standard verification model. *IEEE Trans. Circuits and Systems for Video Technology, 7*(1), 19-31.

Stelmach, L. B., & Tam, W. J. (1994). Processing image sequences based on eye movements. *Proceedings of SPIE Human Vision, Visual Processing, and Digital Display*: *Vol. 2179* (pp. 90-98). San Jose, CA: SPIE.

Viola, P., & Jones, M. (2004). Robust real-time object detection. *International Journal of Computer Vision, 57*(2), 137-154.

VQEG. (2000). Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment. Retrieved from http://www.vqeg.org/

Wandell, B. (1995). *Foundations of vision.* Sunderland, MA: Sinauer Associates.

Winkler, S. (1999). A perceptual distortion metric for digital color video. *Proceedings of SPIE Human Vision and Electronic Imaging*: *Vol. 3644* (pp. 175-184). San Jose, CA: SPIE.

Winkler, S. (2001). Visual fidelity and perceived quality: Towards comprehensive metrics. *Proceedings of SPIE Human Vision and Electronic Imaging*: *Vol. 4299* (pp. 114-125). San Jose, CA: SPIE.

Winkler, S. (2005a). *Digital video quality—Vision models and metrics.* Chichester, UK: John Wiley & Sons.

Winkler, S. (2005b). Video quality metrics—A review. In H. R. Wu & K. R. Rao (Eds.), *Digital video image quality and perceptual coding* (pp. 155-179). Boca Raton, FL: CRC Press.

Winkler, S., & Campos, R. (2003). Video quality evaluation for Internet streaming applications. *Proceedings of SPIE Human Vision and Electronic Imaging*: *Vol. 5007* (pp. 104-115). San Jose, CA: SPIE.

Wolfe, J. M. (1998). Visual search: A review. In H. Pashler (Ed.), *Attention* (pp. 13-74). London: University College London Press.

Wright, M. J., & Johnston, A. (1983). Spatio-temporal contrast sensitivity and visual field locus. *Vision Research, 23*(10), 983-.989.

Yarbus, A. (1967). *Eye movements and vision.* New York: Plenum Press.

ENDNOTES

1. Note that this applies only to spatial contrast sensitivity. The temporal contrast sensitivity does not vary across the visual field (Wright & Johnston, 1983).
2. The no-reference metric used here is part of Genista's quality measurements solutions; see http://www.genista.com for more information.

This work was previously published in Digital Multimedia Perception and Design, edited by G. Ghinea, and S. Y. Chen, pp. 1-20, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.6
A Multidimensional Approach for Describing Video Semantics

Uma Srinivasan
CSIRO ICT Centre, Australia

Surya Nepal
CSIRO ICT Centre, Australia

ABSTRACT

In order to manage large collections of video content, we need appropriate video content models that can facilitate interaction with the content. The important issue for video applications is to accommodate different ways in which a video sequence can function semantically. This requires that the content be described at several levels of abstraction. In this chapter we propose a video metamodel called VIMET and describe an approach to modeling video content such that video content descriptions can be developed incrementally, depending on the application and video genre. We further define a data model to represent video objects and their relationships at several levels of abstraction. With the help of an example, we then illustrate the process of developing a specific application model that develops incremental descriptions of video semantics using our proposed video metamodel (VIMET).

INTRODUCTION

With the convergence of Internet and Multimedia technologies, video content holders have new opportunities to provide novel media products and services, by repurposing the content and delivering it over the Internet. In order to support such applications, we need video content models that allow video sequences to be represented and managed at several levels of semantic abstraction. Modeling video content to support semantic retrieval is a hard task, because *video semantics* means different things to different people. The MPEG-7 Community (ISO/ISE 2001) has spent considerable effort and time in coming to grips with ways to describe video semantics at several levels, in order to support a variety of video applications. The task of developing content models that show the relationships across several levels of video content descriptions has been left to application developers. Our aim in this chapter is to

provide a framework that can be used to develop video semantics for specific applications, without limiting the modeling to any one domain, genre or application.

The Webster Dictionary defines the meaning of semantics as "the study of relationships between "signs and symbols" and what they represent." In a way, from the perspective of feature analysis work (MPEG, 2000; Rui, 1999; Gu, 1998; Flickner et al., 1995; Chang et al. 1997; Smith & Chang, 1997), low- level audiovisual features can be considered as a subset, or a part of, visual "signs and symbols" that convey a meaning. In this context, audio and video analysis techniques have provided a way to model video content using some form of constrained semantics, so that video content can be retrieved at some basic level such as shots. In the larger context of video information systems, it is now clear that feature analyses alone are not adequate to support video applications. Consequently, research focus has shifted to analysing videos to identify higher-level semantic content such as objects and events. More recently, video semantic modeling has been influenced by film theory or semiotics (Hampapur, 1999; Colombo et al., 2001; Bryan-Kinns, 2000), where a meaning is conveyed through a relationship of signs and symbols that are manipulated using editing, lighting, camera movements and other cinematic techniques. Whichever theory or technology one chooses to follow, it is clear that we need a video model that allows us to specify relationships between signs and symbols across video sequences at several levels of interpretation (Srinivasan et al., 2001).

The focus of this chapter is to present an approach to modeling video content, such that video semantics can be described incrementally, based on the application and the video genre. For example, while describing a basketball game, we may wish to describe the game at several levels: the colour and texture of players' uniforms, the segments that had the crowd cheering loudly, the goals scored by a player or a team, a specific movement of a player and so on. In order to facilitate such descriptions, we have developed a framework that is generic and not definitive, but still supports the development of application specific semantics.

The next section provides a background survey of some of these approaches used to model and represent the semantics associated with video content. In the third section, we present our Video Metamodel Framework (VIMET) that helps to model video semantics at different levels of abstraction. It allows users to develop and specify their own semantics, while simultaneously exploiting results of video analysis techniques. In the fourth section, we present a data model that implements the VIMET metamodel. In the fifth section we present an example. Finally, the last section provides some conclusions and future directions.

BACKGROUND

In order to highlight the challenges and issues involved in modeling video semantics, we have organized video semantic modeling approaches into four broad categories and cite a few related works under each category.

Feature Extraction Combined with Clustering or Temporal Grouping

Most approaches that fall under this category have focused on extracting visual features such as colour and motion (Rui et al., 1999; Gu, 1998; Flickner et al., 1995), and abstracting visual signs to identify semantic information such as objects, roles, events and actions. Some approaches have also included audio analysis to extract semantics from videos (MPEG MAAATE, 2000). We recognize two types of features that are extracted from videos: static features and motion-based features. Static features are mostly perceptual features extracted from stationary images or

single frames. Examples are colour histograms, texture maps, and shape polygons (Rui et al., 1999). These features have been exploited as attributes that represent colour, texture and shape of objects and images in a single frame. Motion-based features are those that exploit the basic spatiotemporal nature of videos (Chang et al., 1997). Examples of motion-based low-level features are motion-vectors for visual features and frequency bands for audio signals. These features have been used to segment videos into shots or temporal segments. Once the features are extracted, they are clustered into groups or segmented into temporal units to classify video content at a higher semantic level. Hammoud et al. (2001) use such an approach for modeling video semantics. They extract shots, use clustering to identify similar shots, and use a time-space graph to represent temporal relationships between clusters for extracting scenes as semantic units. Hacid et al. (2000) present a database approach for modeling and querying video data. They use two layers for representing video content: feature and content layer, and semantic layer. The lowest layer is characterized by a set of techniques and algorithms for visual features and relationships. The top layer contains objects of interest, their description and their relationships. A significant component of this approach is the use of temporal cohesion to attach semantics to the video.

Relationship-Based Approach for Modeling Video Semantics

This approach is based on understanding the different types of relationships among video objects and their features. For example, in videos, spatial and temporal relationships of features that occur over space and time have contributed significantly for identifying content at a higher level of abstraction. For example, in the MPEG domain, DCT coefficients and motion vectors are compared over a temporal interval to identify shots and camera operations (Meng et al., 1995). Similarly, spatial relationships of objects over multiple frames are used to identify moving objects (Gu, 1998). Smith and Benitez (2000) present a Multimedia-Extended Entity Relationship (MM-EER) model that captures various types of structural relationships such as spatiotemporal, intersection, and composition, and provide a unified multimedia description framework for content-based description of multimedia in large databases. The modeling framework supports different types of relationships such as generalization, aggregation, association, structural relationship, and intensional relationship. Baral et al. (1998) also extended ER diagrams with a special attribute called "core." This attribute stores the real object in contrast to the abstraction which is reflected in the rest of the attributes. Pradhan et al. (2001) propose a set of interval operations that can be used to build an answer interval to the given query by synthesizing video interval containing keywords similar to that of the query.

Theme-Based Annotation Synchronized to Temporal Structure

This approach is based on using textual annotations synchronized to a temporal structure (Yap et al. 1996). This approach is often used in television archives where segments are described thematically and linked to temporal structures such as shots or scenes. Hjelsvold and Midtstraum (1994) present a general data model for sharing and reuse of video materials. Their data model is based on two basic ideas: temporal structural representation of video and thematic annotations and relationships between them. Temporal structure of video includes frame, shot, scene, sequence, and so forth. The data model allows users to describe frame sequences using thematic annotations. It allows detailed descriptions of the content of the video material, which are not necessarily linked to structural components. More often they are linked to arbitrary frame sequences by establishing a relationship between the frame sequences

and relevant annotations, which are independent of any structural components. Other works that fall under this category include (Bryan-Kinns, 2000; Srinivasan, 1999).

Approach Based on Knowledge Models and Film Semiotics

As we move up the semantic hierarchy to represent higher-level semantics in videos, in addition to extracting features and establishing their relationships, we need knowledge models that help us understand higher-level concepts that are specific to a video genre. Models that aim to represent content at higher semantic levels use some form of knowledge models or semiotic and film theory to understand the semantics of videos. Film theory or semiotics describes how a meaning is conveyed through a relationship of signs and symbols that are meaningful to a particular culture. This is achieved by manipulation of editing, lighting, camera movements and other cinematic techniques.

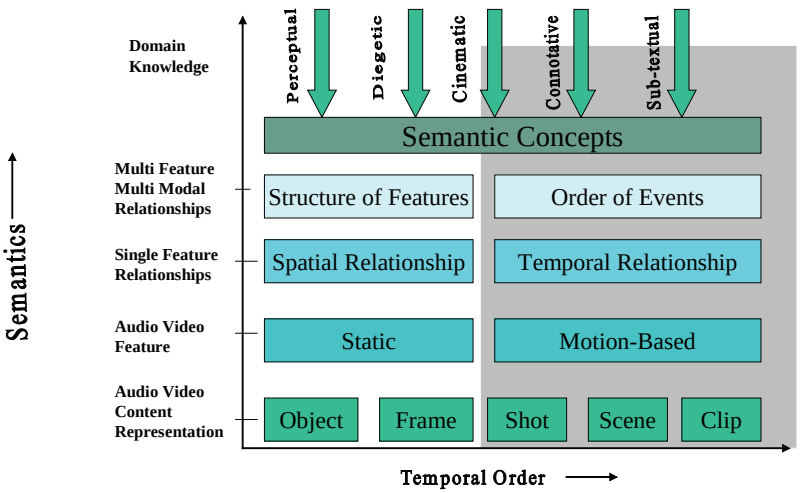
Colombo et al. (2001) present a model that allows retrieval of commercial content based on their salient semantics. The semantics are defined from the semiotic perspective, that is, collections of signs and semantic features like colour, motion, and so forth, are used to build semantics. It classifies commercials into four different categories based on semiotics: practical, critical, utopic and playful. A set of rules is then defined to map the set of perceptual features to each of the four semiotic categories for commercials.

Bryan-Kinns (2000) presents a framework (VCMF) for video content modeling which represents the structural and semantic regularities of classes of video recordings. VCMF can incorporate many domain-specific knowledge models to index video at the semantic level.

Hampapur (1999) presents an approach based on the use of knowledge models to build domain-specific video information systems. His aim is to provide higher-level semantic access to video data. The knowledge model proposed in their paper includes the semantic model, cinematic model and physical world model. Each model has several sub-models; for example, the cinematic model includes a camera motion model and a camera geometry model.

We realize that the above categories are not mutually exclusive and many models and frameworks use a combination of two or more of these modeling strategies.

Figure 1. A multidimensional data model for building semantics (the shadow part represents the temporal (dynamic) component)



We now describe our metamodel framework. It uses feature extraction, multimodal relationships, and semiotic theory to model video semantics at several levels of abstraction. In a way, we use a combination of the four approaches described above.

VIDEO METAMODEL (VIMET)

Figure 1 shows the video metamodel which we call VIMET. The horizontal axis shows the temporal dimension of the video and the vertical axis shows the semantic dimension. A third dimension represents the domain knowledge associated with specialized interpretations of a film or a video (Metz, 1974; Lindley & Srininvasan, 1999). The metamodel represents: (1) static and motion-based features in both audio and visual domain, (2) different types of spatial and temporal relationships, (3) multimodal, multifeature relationships to identify events, and (4) concepts influenced by principles of film theory. The terms and concepts used in the metamodel are described below.

The bottom layer shows the standard way in which a video is represented to show temporal granularity.

- **Object:** It represents a single object within a frame. An object may be identified manually or automatically if an appropriate object detection algorithm exists.
- **Frame:** It represents a single image of a video sequence. There are 25 frames (or 30 in certain video formats) in a video of one-second duration.
- Shot: A shot is a temporal video unit between two distinct camera operations. Shot boundaries occur due to different types of camera operations such as cuts, dissolves, fades and wipes. There are various techniques used to determine shots. For example, DCT coefficients are used in MPEG videos to identify distinct shot boundaries.
- **Scene:** A scene is a series of consecutive shots constituting a unit from the narrative point of view sharing some thematic visual content. For example, a video sequence that shows Martina Hinges playing the "game point" in the 1997 Australian open final is a thematic scene.
- **Clip:** A clip is an arbitrary length of video used for a specific purpose. For example, a video sequence that shows tourist places around Sydney is a video clip made for tourism purpose.

The semantic dimension helps model semantics in an incremental way using features and their relationships. It uses information from the temporal dimension to develop the semantics.

- **Static features:** Represent features extracted from objects and frames. Shape, color and texture are examples of static features extracted from the objects. Similarly, global color histogram and average colors are examples of features extracted from a frame.
- **Motion-based features:** Represent features extracted from video using motion-based information. An example of such a feature is motion vector.

The next level of conceptualisation occurs when we group individual features to identify objects and events using some criteria based on spatial and/or temporal relationships.

- **Spatial Relationships:** Two types of spatial relationships are possible for videos: topological relationships and directional relationships. Topological relationships include relations such as contains, covered by, and disjoint (Egenhofer & Franzosa, 1991). Directional relationships include relations such as right-of, left-of, above and below (Frank, 1996).

- **Temporal Relationships:** This includes the most commonly used temporal relationship from Allen's 13 temporal relationships (Allen, 1983). These include temporal relations such as before, meets, overlaps, finishes, starts, contains, equals, during, started by, finished by, overlapped by, met by and after.

Moving further up in the semantic dimension, we use multiple features both in the audio and the visual domain, and establish (spatial and temporal) relationships across these multimodal features to model higher-level concepts. In the temporal dimension, this helps us identify semantic constructs based on structure of features and order of events occurring in a video.

- **Structure of Features:** Represents patterns of features that define a semantic construct. For example, a group of regions connected through some spatial relationship over multiple frames, combined with some loudness value spread over a temporal segment, could give an indication of a particular "event" occurring in the video.
- **Order of Events:** Represents recurring pattern of events identified manually or detected automatically using multimodal feature relationships. For example, camera pan is a cinematic event that is derived by using a temporal relationship of motion vectors. (The motion vectors between consecutive frames due to the pan should point in a single direction that exhibits a strong modal value corresponding to the camera movement.) Similarly, a sudden burst of sound is a perceptual auditory event. These cinematic and perceptual events may be arranged in certain sequence from a narrative point of view to convey a meaning. For example, a close-up followed by a loud sound could be used to produce a dramatic effect. Similarly, a crowd cheer followed by scorecard could be used to determine field goals in basketball videos. (We will describe this in greater details later in our example.)

Modeling "the meaning" of a video, shot, or sequence requires the description of the video object at several levels of interpretation. Film semiotics, pioneered by the film theorist Christian Metz (1974), has identified five levels of cinematic codification that cover visual features, objects, actions and events depicted in images together with other aspects of the meaning of the images. These levels are represented in the third dimension of the metamodel. The different levels interact together, influencing the domain knowledge associated with a video.

- **Perceptual level:** This is the level at which visual phenomena become perceptually meaningful, the level at which distinctions are perceived by the viewer. This is the level that is concerned with features such as colour, loudness and texture.
- **Cinematic level:** This level is concerned with formal film and video editing techniques that are incorporated to produce expressive artifacts. For example, arranging a certain rhythmic pattern of shots to produce a climax, or introducing voice-over to shift the gaze.
- **Diegetic level:** This refers to the four-dimensional spatiotemporal world posited by a video image or a sequence of video images, including spatiotemporal descriptions of objects, actions, or events that occur within that world.
- **Connotative level:** This level of video semantics is the level of metaphorical, analogical and associative meanings that the objects and events in a video may have. An example of connotative significance is the use of facial expression to denote some emotion.

- **Subtextual level:** This is the level of more specialized, hidden and suppressed meanings of symbols and signifiers that are related to special cultural and social groups.

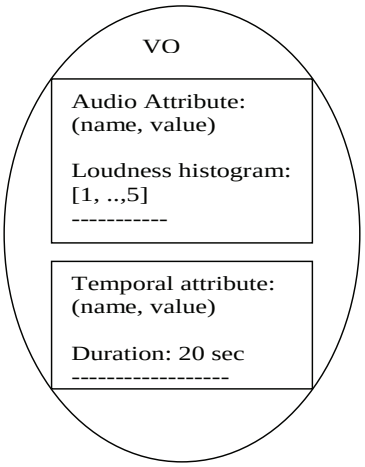
The main idea of this metamodel framework is to allow users to develop their own application models, based on their semantic notion and interpretation, by specifying objects and relationships of interest at any level of granularity.

Next we describe the data model to implement the ideas presented in the VIMET metamodel. The elements of the data model allow application (developers) to incrementally develop video semantics that needs to be modeled and represented in the context of the application domain.

VIDEO DATA MODEL

The main elements of the data model are: (i) Video Object (VO) to model a video sequence of any duration, (ii) VO Attributes that are either explicitly specified or based on automatically detected audio, visual, spatial and temporal features, (iii) VO Attribute-level relationships for computed attribute values, (iv) Video Concept Object (VCO) to accommodate fuzzy descriptions, and (v) Object level relationships, to support multimodal and multifeature relationships across video sequences.

Figure 2. A typical video object with two sets of attributes: audio and temporal



Video Object (VO)

We define a video object (VO) as an abstract object that models a video sequence at any level of abstraction in the semantic-temporal dimension shown in Figure 1. This is the most primitive object that can be used in a query.

Definition: A typical Video Object (VO) is a five-tuple

VO = <Xf, Sf, Tf, Vf , Af)

where

Xf is a set of textual attributes,
Sf is a set of spatial attributes,
Tf is a set of temporal attributes,
Vf is a set of visual attributes, and
Af is a set of audio attributes.

Xf represents a set of textual attributes that describe the semantics associated with the object at different levels of abstraction. This could be metadata or any textual description — a manual annotation of the content or about the content.

Sf represents spatial attributes that specify the spatial bounds of the object. This pertains to the space occupied by the object in a two-dimensional plane. This could be X and Y positions of the object or a bounding box of the object.

Ti represents the temporal attributes to describe the temporal bounds of the object. Temporal attributes includes start time, end time, and so forth. A time interval is a basic primitive used here.

Vf represents a set of attributes that characterize the object in terms of visual features. The values of these attributes are typically the result of visual feature extraction algorithms. Examples

of such attributes/features are colour histograms and motion vectors.

Af represents a set of attributes that characterize the object in terms of aural features. The values of these attributes are typically the result of audio analysis and feature extraction algorithms. Examples of such attributes/features are loudness curves and pitch values.

A typical video object — a video shot — is shown in Figure 2. The diagram shows a shot with temporal attribute duration and audio attribute loudness histogram. (To keep the diagram simple we have shown only a subset of possible attributes of a video object.) Each attribute has a value domain and is shown as a name value pair. For example, the audio attribute set has an audio attribute loudness whose value is given by a loudness histogram. Similarly, the temporal attribute set has an attribute duration whose value is given in seconds.

The applications determine the content of the video objects modeled. For example, a video object could be a frame in the temporal dimension, with an attribute specifying its location in the video. In the semantic dimension, the same frame could have visual features such as colour and texture histograms.

VO Attributes

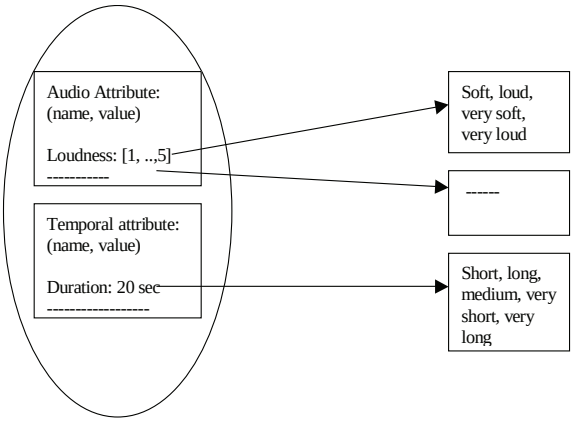
The attributes of a video object are either intensional or extensional. Extensional attribute values are text data, drawn from the character domain. The possible sources of extensional attributes are annotation, transcripts, keywords, textual description, terms from a thesaurus, and so forth. Extensional attributes fall in the feature category in Figure 1. Intensional attributes have specific value domains where the values are computed using appropriate feature extraction functions. Where the extensional attributes of the video objects are semantic in nature, relationships across such objects can be expressed as association, aggregation, and generalisation as per object-oriented or EER modeling methodologies. For intensional attributes, however, we need to establish specific relationships for each attribute type. For example, when we consider the temporal attribute whose value domain is "time-interval," we need to specify temporal relationships and corresponding operations that are valid over time intervals.

We define two sets of value domains for intensional attributes. The first set is the numerical domain and the second set is the linguistic domain. The purpose of providing a linguistic domain is to allow a fuzzy description of some attributes. For example, a temporal attribute duration could have a value equal to 20 (seconds) in the numerical domain. For the same temporal attribute, we define a second domain, which is a constrained set of linguistic terms such as "short," "long" and "very long." For this, we use a fuzzy linguistic framework, where the numerical values of duration are mapped to fuzzy terms such as short, long, and so forth. (Driankov et al., 1993). The fuzzy linguistic framework is defined as a four tuple <X, LX, QX, MX> where

- X = denotes the symbolic name of a linguistic variable (attributes in our case) such as duration.
- LX = set of linguistic terms that X can take such as "short," "long" and "very long.".
- QX = the numeric domain where X can take the values such as "time in seconds."
- MX = a semantic function which gives "meaning" to the linguistic terms by mapping X's values from LX to QX. The function MX depends on the types of attributes and the application domain. We defined one such function in Nepal et al. (2001).

Similarly, when we consider an audio attribute value based on the loudness histogram/curve, we map the numerical values to semantic terms such as "loud,", "soft," "very loud" and so on. By using a fuzzy linguistic framework, we are utilising

Figure 3. Mapping of numerical attribute values to fuzzy attribute values



the ability of the human to distinguish variations in audio and visual features. Figure 3 shows the mapping of a numerical attribute-value domain to a linguistic attribute-value domain.

VO Attribute-Level Relationships

Relationships across video objects can be defined at two levels: the object and attribute levels. We will discuss the object level relationships in a later section. Here we discuss attribute-level relationships to illustrate the capabilities of the data model. Each intensional attribute value allows us to define relationships that are valid for that particular attribute-value domain. For example, when we consider visual relationships based on the "brightness" attribute, we should be able to establish relationships such as "brighter-than," "dimmer-than," and "similar" by grouping light intensity values. By exploiting such attribute-level relationships across each intensional attribute, it is possible to establish a multidimensional relationship between video objects. Here each dimension reflects a perceivable variation in the computed values of the relevant intensional attribute.

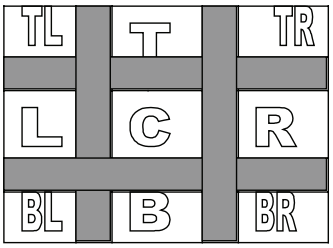
Table 1 shows an illustrative set of attribute-level relationships. The relationship values are drawn from a predefined constrained set of relationships valid for a particular attribute type. A typical relationship value set for brightness is {brighter-than, dimmer-than, similar}. Each relationship in the set is given a threshold based on the eye's ability to discriminate between light intensity levels. The fuzzy scale helps in accommodating subjectivity at the user level.

Next we give some examples of directional relationships used for spatial reasoning in large-scale spaces, that is, spaces that cannot be seen or understood from a single point of view. Such relationships are useful in application areas such as Geographical Information Systems. In the case of video databases, it is more meaningful to define equivalent positional relationships as we are dealing with spaces within a frame that can be viewed from a point. The nine positional relationships are given by AP = {Top, TopLeft, Left, BottomLeft, Bottom, BottomRight, Right, TopRight, Centre} as shown in Figure 4. These nine relationships can be used as linguistic values for a spatial attribute "position" given by a bounding box. Similarly, appropriate linguistic terms can be developed for other attributes as shown in Table 1.

Table 1. Attribute-level relationships

Attribute	*Relationship Type*	*Relationship Value set*
Brightness	Visual, based on light intensity	{Brighter, Dimmer, Similar}
Contrast	Visual, based on contrast measures	{Higher, Lower, Similar}
Color	Visual, based on color	(Same, Different, Resembles}
Loudness	Audio, based on sound level	{Louder, Softer, Similar}
Pitch	Audio, based on pitch	{Higher, Lower, Similar}
Size	Visual, based on size	{Bigger, Smaller, Similar}
Duration	Temporal, based on duration	{Shorter, Longer, Equal}

Figure 4. Positional relations in a two-dimensional space



Video Concept Object

In order to support the specification of an attribute relationship (using linguistic terms), we define a video concept object. A Video Concept Object (VCO) is a video object with a semantic attribute value attached to each intensional attribute. The main difference between a VO and a VCO lies in the domain of the intensional attributes. The attributes of a VO have numerical domains, and the attributes of VCO have a domain which is a set of linguistic terms which forms a domain for semantic attribute value. The members of this set are controlled vocabulary terms that reflect the (fuzzy) semantic value for each attribute type. The domain set will be different for audio, visual, temporal and spatial attributes.

Definition: A typical Video Concept Object (VCO) is a five-tuple

VCO = <Xc, Sc, Tc, Ac, Vc>

where

Xc is a set of textual attributes that define a concept, the attribute values are drawn from the character domain,

Sc represents a set of spatial attributes. The value domain is a set of (fuzzy) terms that describe relative positional and directional relationships,

Tc represents a set of temporal attributes whose values are drawn from a set of fuzzy terms used to describe time interval or duration.

Ac represents a set of audio attributes (an example is loudness attribute whose values are drawn from a set of fuzzy terms that describe loudness).

Vc represents a visual attribute whose values are drawn from a set of fuzzy terms that describe that particular visual attribute.

The relationship between a VO and a VCO is established using fuzzy linguistic mapping as shown in Figure 5.

In a fuzzy linguistic model, attribute name-relationship pair is equivalent to fuzzy variable-value pair. In general, a user can query and retrieve video from a database using any primitive VO/VCO. An example is "retrieve all videos where an object A is at the right-bottom of the frame." In this example, the expressive power of a query is based on a single attribute-value and is limited. A more interesting query would be "retrieve all video sequences where the loudness value is very high and the duration of the sequence is short." Such queries would include VCOs and VOs.

VO and VCO Relationships

In order to retrieve sequences with multiple relationships over different VOs and VCOs, we need appropriate operators. In addition, we need some kind of knowledge or heuristic rule(s) that help us interpret these relationships. This motivates us to define a Video Semantics System.

Definition: A typical Video Semantics System (VSS) is a five-tuple

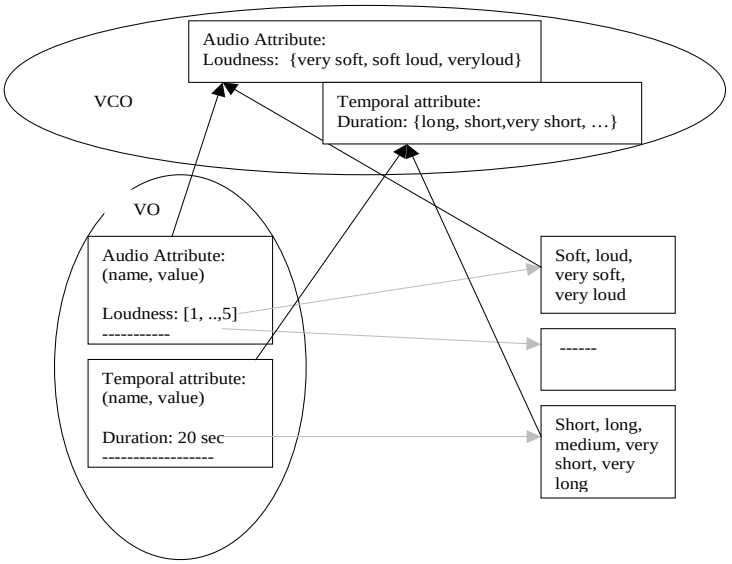
VSS = < VO/VCO, Sro, Tro, Fro, I>

where

VO/VCO is a set of Video Objects or Video Concept Objects,

Sro is a set of spatial relationship operators,

Figure 5. An example VCO derived using a fuzzy linguistic model



Tro is a set of temporal relationship operators, and
Fro is a set of multimodal operators.
I is an interpretation model of the video object.

Next we describe some operators and interpretation models. We define spatial operators for manipulating relationships of objects over spatial attributes. This helps in describing the structure of features (Figure 1). We define temporal operators for manipulating relationships over temporal attributes. This helps to describe order of events (Figure 1). For multimodal multifeature composition, however, we define fuzzy Boolean operators.

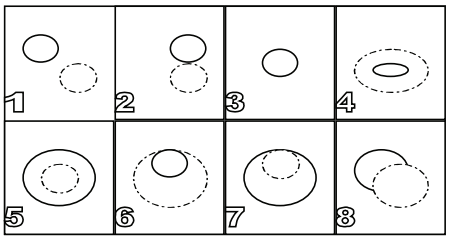
Spatial Relationship Operators (SROs)

A number of spatial relationships with corresponding operators have been proposed and used in spatial databases, geographical information systems and multimedia database systems. We describe them into the following categories. It is important to note that SROs are valid only if the VO has attributes (values) that indicate spatial bounds.

- **Topological Relations:** Topological relations are spatial relations that are invariant under bijective and continuous transformations that have also continuous inverses. Topological equivalence does not necessarily preserve distances and directions. Instead, topological notions include continuity, interior, and boundary, which are defined in terms of neighbourhood relations. The eight topological relationship operators are given as TR = {Disjoint, Meet, Equal, Inside, Contains, Covered_By, Covers, Overlap} (Egenhofer & Franzosa, 1991) as shown in Figure 6. These relationship operators are binary.
- **Relative Positional Relations:** The relative positional relationship operators are given by RP = {Right of, Left of, Above, Below, In front of, Behind}. These operators are used to express the relationships among VCOs. These relative positional relationship operators are binary.

Sro = TR ∪ RP

Figure 6. Examples of eight basic topological relationships



Temporal Relationship Operators (TROs)

As a video is a continuous medium, temporal relations provide important cues for video data retrieval. Allen (1983) has defined 13 temporal interval relations. Many variations of Allen's temporal interval relations have been proposed and used in temporal and multimedia databases. These relations are used to define the temporal relationships of events within a video. For example, weather news appears after the sports news in a daily TV news broadcast. It is important to note that TROs are valid only for VOs that have attributes with temporal bounds. The 13 temporal relationships defined by Allen are given by TR = {Before, Meets, Overlaps, Finishes, Starts, Contains, Equals, During, Started by, Finished by, Overlapped by, Met by, After}. (Note: This set TR corresponds to the set of temporal relationship operators defined as Tro in VSS.)

Multimodal Operators

Traditionally the approach used for audiovisual content retrieval is based on similarity measures on the extracted features. In such systems, users need to be familiar with the underlying features and express their queries in terms of these (low-level) features. In order to allow users to formulate semantically expressive queries, we define a set of fuzzy terms that can be used to describe a feature using fuzzy attribute relationships. For example, when we are dealing with the "loudness" feature, we use a set of fuzzy terms such as "very loud," "loud," "soft," and "very soft" to describe the relative loudness at the attribute level. This fuzzy measure is described at the VCO level. A query on VSS, however, can be multimodal in nature and involve many VCOs. In order to have a generic way to combine multimodal relationships, we use simple fuzzy Boolean operators. The fuzzy Boolean operator set is given by

FR = {AND, OR, NOT}.

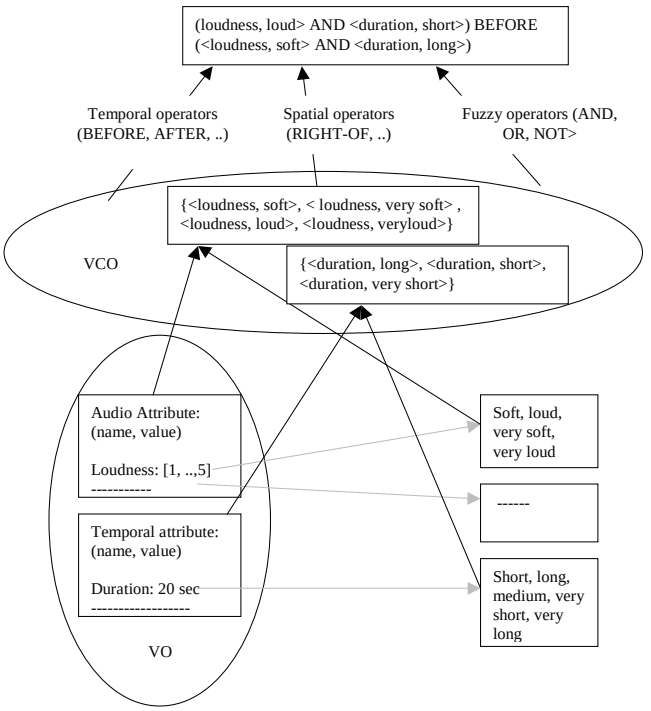
(Note: This set FR corresponds to the set of multimodal relationship operators defined as Fro in VSS.)

An example of a multimodal, multifeature query is shown in Figure 7.

Interpretation Models

In the third section we identified five different levels of cinematic codification and description that help us interpret the "meaning" of a video. Lindley and Srinivasan (1998) have demonstrated empirically that these descriptions are meaningful in capturing distinctions in the way images are viewed and interpreted by a nonspecialist audience, and between this audience and the analytical terms used by filmmakers and film critics. Modeling "the meaning" of a video, shot, or sequence requires the description of the video object at any or all of the interpretation levels described in the third section. Interpretation Models can be based on one or several of these five levels. A model based on perceptual visual characteristics is the subject of a large amount of current research on video content-based retrieval (Ahanger & Little, 1996). Models based on cinematic constructs incorporate the expressive artifacts such as camera operations (Adam et al., 2000), lighting schemes and optical effects (Truong et al., 2001). Automated detection of cinematic features is another area of vigorous current research activity (see

Figure 7. An example of multimodal query using fuzzy and temporal operators on VCOs



Ahanger & Little). While modeling at the diegetic level the basic perceptual features of an image are organised into a four-dimensional spatiotemporal world posited by a video image or sequence of video images. This includes the spatiotemporal descriptions of agents, objects, actions and events that take place within that world. The interpretation model that we illustrate in the fifth section is a diegetic model based on interpretations at the perceptual level. Examples of connotative meanings are the emotions connoted by actions or the expressions on the faces of characters. The subtextual level of interpretation involves representing specialised meanings of symbols and signifiers. For both the connotative and the subtextual levels, definitive representation of the "meaning" of a video is in principle impossible. The most that can be expected is an evolving body of interpretations.

In the next section we show how a user/author can develop application semantics by explicitly specifying objects and relationships and interpretation models. The ability to create different models allows users/authors to contextualise the content to be modeled.

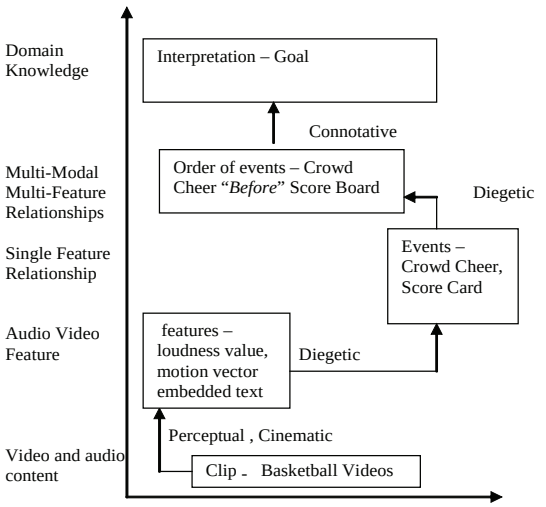
EXAMPLE

Here we illustrate the process of developing a specific application model to describe video content at different semantic levels. The process is incremental, where the different elements of the VIMET framework are used to retrieve "goal" segments from basketball videos. Here we have used a top-down approach to modeling, where we first developed the interpretation models from an empirical investigation of basketball videos (Nepal et al., 2001). We then develop an object model for the application.

Figure 8 shows an instantiation of the metamodel.

The observation of television broadcasts of basketball games gave some insights into com-

Figure 8. An instance of metamodel for a video semantic "Goal" in basketball videos using temporal interpretation model I



monly occurring patterns of events perceived during the course of a basketball game. We have considered a subset of these as key events that occur repeatedly throughout the video of the game to develop a temporal interpretation model. They were: crowd cheer, scorecard display and change in players' direction. We then identified the audio-video features that correspond to the key events — manually observed. The features identified include energy levels of audio signals, embedded text regions, and change in direction of motion. Since the focus of this chapter is on building a video semantic "goal," we will not elaborate on the feature extraction algorithms in detail. We used the MPEGMaaate audio analysis toolkit (MPEG MAAATE, 2000) to identify "high-energy" segments from loudness values. Appropriate thresholds were applied to these high-energy segments to capture an event "crowd cheer." The mapping to crowd cheer expresses a simple domain heuristic (diegetic and connotative interpretation) for sports video content. Visual feature extraction consists of identifying embedded text (Gu, 1998) and is mapped to score card displays in the sports domain. Similarly, camera pan (Srinivasan et al., 1997) motion is extracted and used to capture change in players' direction. The next level of semantics is developed by exploring the temporal order of events such as crowd cheer and scorecard displays. Here we use temporal interpretation models that show relationships on multiple features from multiple modalities.

Interpretation Models

We observed basketball videos and developed five different temporal interpretation models, as follows.

Model I

This model is based on the first key event — crowd cheer. Our observation shows that there is a loud cheer within three seconds of scoring a legal goal. Hence, in this model, the basic diegetic interpretation is that a loud cheer follows every legal goal, and a loud cheer only occurs after a legal goal. The model T1 is represented by

T1: Goal $\rightarrow$ [3 sec] $\rightarrow$ crowd cheer

Intuitively, one can see that the converse may not always be true, as there may be other cases where a loud cheer occurs, for example, when a streaker runs across the field. Such limitations are addressed to some extent in the subsequent models.

Model II

This model is based on the second key event— scoreboard display. Our observation shows that the scoreboard display is updated after each goal. Our cinematic interpretation in this model is that a scoreboard display appears (usually as embedded text) within 10 seconds of scoring a legal goal. This is represented by the model T2.

T2: Goal → [10 sec] → Scoreboard

The limitation of this model is that the converse here may not always be true, that is, a scoreboard display may not always be preceded by a legal goal.

Model III

This model uses a combination of two key events with a view to address the limitations of T1 and T2. As pointed out earlier, all crowd cheers and scoreboard displays may not always indicate a legal goal. Ideally when we classify segments that show a shooter scoring goals, we need to avoid inclusion of events that do not show a goal, even though there may be a loud cheer. In this model, this is achieved by temporally combining the scoreboard display with crowd cheer. Here, our diegetic interpretation is that every goal is followed by crowd cheer within three seconds, and by a scoreboard display within seven seconds after the crowd cheer. This discards events that have crowd cheer, but no scoreboard and events that have scoreboards, but no crowd cheer.

T3: Goal → [3 sec] → Audio Cheer → [7 sec] → Score Board

Model IV

This model addresses the strict constraints imposed in Model 3. Our observations show that while the pattern shown in three is valid most of the times, there are cases where the field goals are accompanied by loud cheers and no scoreboard display. Similarly, there are cases where goals are followed by scoreboard displays but not crowd cheer, as in the case of free throws. In order to capture such scenarios, we have used a combination of models I and II and proposed a model IV.

T4: T1 → T2

where ∪ is the union of results from models H1 and H2.

Model V

While model IV covers most cases of legal goals, due to the inherent limitations pointed out in models I and II, model IV could potentially classify segments where there are no goals. Our observations show that Model IV captures the maximum number of goals, but it also identifies many nongoal segments. In order to retain the number of goal segments and still remove the nongoal segments, we introduce the third key event — change in direction of players. In this model, if a crowd cheer appears within 10 seconds of a change in direction, or a scoreboard appears within 10 seconds of a change in direction, there is likely to be a goal within 10 seconds of the change in direction. This is represented as follows.

T5: Goal → [10 secs] → Change in direction → [10secs] → Crowd cheer

OR

T5:Goal → [10 secs] → Change in direction → [10secs] → Scoreboard

Although, in all models, the time interval between two key-events is hardwired, the main idea is to provide a temporal link between key-events to build up high-level semantics.

Object Model

For the interpretation models described above, we combine video objects characterized by different types of intensional attributes to develop new video objects or video concept objects. An instance of video data model is shown in Figure 9. For example, we use the "loudness" feature and a temporal feature to develop a new video concept object called "crowd cheer". For example, we first define video semantics crowd cheer as.

Figure 9. An instance of the data model for an example explained in this section.

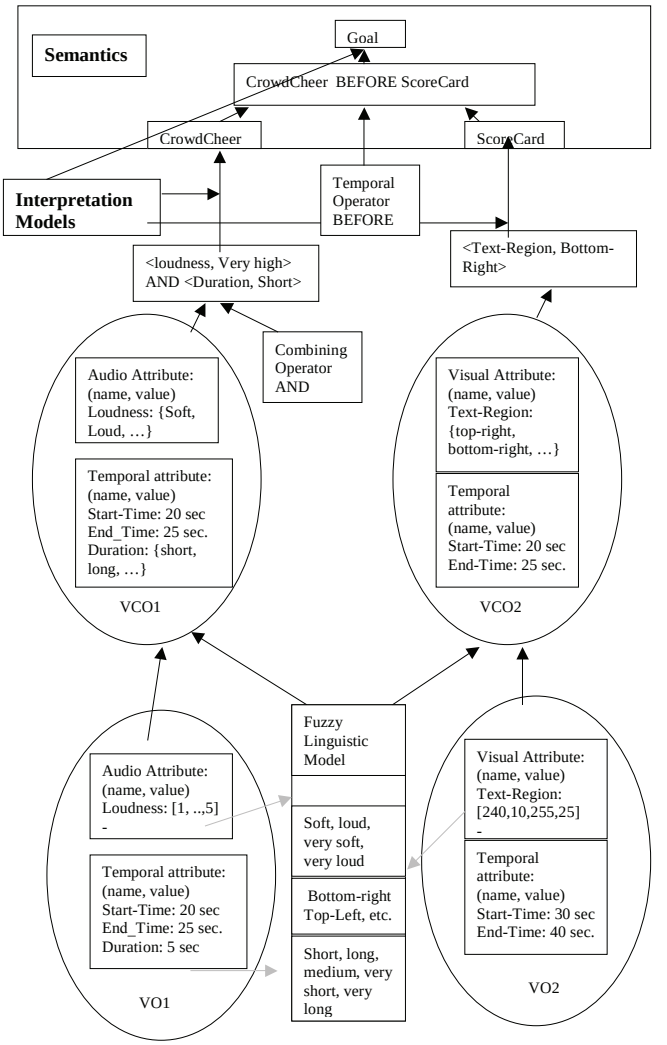


Table 2. A list of data sets used for our experiments and the results of our observations

Games	*Description*	*Length*	*Number of Cheers*	*Number of Scoreboard Displays*	*Number of Goals*	*Number ofchanges in direction of players' movements*
A	Australia Vs Cuba 1996 (Women)	00:42:00	31	49	52	74
B	Australia Vs USA 1994 (Women)	00:30:00	27	46	30	46
C	Australia Vs Cuba 1996 (Women)	00:14:51	16	16	17	51
D	Australia Vs USA 1997 (Men)	00:09:37	13	16	18	48

Table 3. A summary of results of automatic evaluation of various algorithms

Clips	Total number of baskets (relevant)	Algorithms	Total (retrieved)	Correct Decision (relevant ∩ retrieved)	Precision (%)	Recall (%)
C	17	T1	15	12	80	70.50
		T2	20	15	75	88.23
		T3	11	11	100	64.70
		T4	24	16	66.66	94.1
		T5	17	15	88.23	88.23
D	18	T1	16	11	68.75	61.11
		T2	19	17	89.47	94.44
		T3	10	10	100	55.55
		T4	25	18	72.0	100
		T5	22	16	72.72	88.88

```
DEFINE CrowdCheer AS
SELECT V1.Start-time, V1.End-time
FROM VCO V1
WHERE <V1.Loudness, Very-high> AND
        <V1.Duration, Short>
```

Here we use diegetic knowledge about the domain to interpret a segment where a very high loud value appeared for a short duration as a crowd cheer. Similarly, we can define scorecard as follows.

```
DEFINE ScoreCard AS
SELECT V2.Start-time, V2.End-time
FROM VCO V2
WHERE  <V2.Text-Region, Bottom-Right>
```

We then use the temporal interpretation model to build a query to develop a video concept object that represents a "goal" segment. This is defined as follows:

```
DEFINE Goal AS
SELECT [10]  MIN(V1.Start-time, V2.Start-
time),
        MAX(V1.End-time, V2.End-time)
FROM CrowdCheer V1, ScoreCard V2
WHERE V1 BEFORE V2
```

Here the query returns the 10 best-fit video sequences that satisfy the query criteria - goal.

We have implemented this application model for basketball videos. Implementing such a model necessarily involves developing appropriate interpretation models, which is a labour-intensive task. The VIMET framework helps this process by facilitating incremental descriptions of video content.

We now present an evaluation of the interpretation models outlined in the previous section.

Evaluation of Interpretation Models

The data set used to evaluate the different temporal models for building a video semantic "goal" in basketball videos is shown in Table 2. The first two clips A and B are used for evaluating observations and the last two clips are used for evaluating automatic analysis.

The standard precision-recall method was used to compare our automatically generated

goal segments with the ones manually judged by humans.

Precision is the ratio of the number of relevant clips retrieved to the total number of clips retrieved. The ideal situation corresponds to 100% precision, when all retrieved clips are relevant.

$$precision = \frac{|\ relevant \cap retrieved\ |}{|\ retrieved\ |}$$

Recall is the ratio of the number of relevant clips retrieved to the total number of relevant clips. We can achieve ideal recall (100%) by retrieving all clips from the data set, but the corresponding precision will be poor.

$$recall = \frac{|\ relevant \cap retrieved|}{|\ relevant|}$$

We evaluated the temporal interpretation models in our example data set. The manually identified "legal goals" (which include field goals and free throws) of the videos in our data set are shown in column 2 in Table 3.

The result of automatic analysis shows that the combination of all three key events performs much better with high recall and precision values (~88%). In model T1, the model correctly identifies 12 goal segments out of 17 for video C and 11 out of 18 for video D. However, the total number of crowd segments detected by our crowd cheer detection algorithm is15 and 16 for videos C and D, respectively. That is, there are few legal goal segments that do not have crowd cheer and vice versa. Further analysis shows that crowd cheer events resulting from other interesting events such as "fast break", "clever steal" and "great check or screen" gives false positive results. We observed that most of the field goals are accompanied by crowd cheer. However, many goals scored by free throws are not accompanied by crowd cheer. We also observed that in certain cases lack of supporters among the spectators for a team yield false negative results. In model T2, the model correctly identifies 15 goals out of 17 in video C and 17 out of 18 in video D. Our further analysis confirmed that legal goals due to free throws are not often accompanied by scoreboard displays, particularly when the scoreboard is updated at the end of free throws rather than after each free throw. Similarly, our feature extraction algorithm used for scoreboard not only detects scoreboards but also other textual displays such as team coach names and the number of fouls committed by a player. Such textual features increase the number of false positive results. We plan to use the heuristics developed here to improve our scoreboard detection algorithm in the future. The above discussion is valid for algorithms T3, T4 and T5 as well.

CONCLUDING REMARKS

Emerging new technologies in video delivery such as streaming over the Internet have made video content a significant component of the information space. Video content holders now have an opportunity to provide new video products and services by reusing their video collections. This requires more than content-based analysis systems.

We need effective ways to model, represent and describe video content at several semantic levels that are meaningful to users of video content. At the lowest level, content can be described using low-level features such as colour and texture, and at the highest-level the same content can be described using high-level concepts. In this chapter, we have provided a systematic way of developing content descriptions at several semantic levels. In the example in the fifth section, we have shown how (audio and visual) feature extraction techniques can be effectively used together with interpretation models, to develop higher-level semantic descriptions of content. Although the example is specific to the sports genre, the VIMET framework and the associated data model provides a platform to develop several descriptions of the same video content, depending on the interpreta-

tion level of the annotator. The model supports one of the mandates of MPEG-7 by supporting content descriptions, to accommodate multiple interpretations.

REFERENCES

Adam, B., Dorai, C., & Venkatesh, S. (2000). Study of shot length and motion as contributing factors to movie tempo. *ACM Multimedia*, 353-355.

Ahanger, G., & Little, T.D.C. (1996). A survey of technologies for parsing and indexing digital video. *Journal of Visual Communication and Image Representation*, *7*(1), 28-43.

Allen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communication of the ACM*, *26*(11), 832-843.

Baral, C., Gonzalez, G., & Son, T. (1998). Conceptual modeling and querying in multimedia databases. *Multimedia Tools and Applications*, *7*(½), 37-66.

Bryan-Kinns, N. (2000). VCMF: A framework for video content modeling. *Multimedia Tools and Applications*, *10*(1), 23-45.

Chang, S.F., Chen, W., Meng, H.J., Sundaram, H., & Zhong, D. (1997). VideoQ: An automated content based video search system using visual cues. *ACM Multimedia*, 313-324, Seattle, Washington, November.

Colombo, C., Bimbo, A.D., & Pala, P. (2001). Retrieval of commercials by semantic content: the semiotic perspective. *Multimedia Tools and Applications*, *13*, 93-118.

Driankov, D., Hellendoorn, H., & Reinfrank, M. (1993). *An introduction to fuzzy control*. Springer-Verlag.

Egenhofer, M., & Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographic Information Systems*, *5*(2), 161-174.

Fagin, R. (1999). Combining fuzzy information from multiple systems. *Proceedings of the 15th ACM Symposium on Principles of Database Systems* (pp. 83-99).

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., & Steele, D. (1995). Query by image and video content: The QBIC system. *Computer*, *28*(9), 23-32.

Frank, A.U. (1996). Qualitative spatial reasoning: Cardinal directions as an example. *International Journal of Geographic Information Systems*, *10*(3), 269-290.

Gu, L. (1998). Scene analysis of video sequences in the MPEG Domain. *Proceedings of the IASTED International Conference Signal and Image Processing*, October 28-31, Las Vegas.

Hacid, M.S., Decleir, C., & Kouloumdjian, J. (2000). A database approach for modeling and querying video data. *IEEE Transaction on Knowledge and Data Engineering*, *12*(5), 729-750.

Hammoud, R., Chen, L., & Fontaine, D (2001). An extensible spatial-temporal model for semantic video segmentation. *TRANSDOC project*. Online at *http:://transdoc.ibp.fr/*

Hampapur, A. (1999). Semantic video indexing: Approaches and issues. *SIGMOD Record*, *28*(1), 32-39.

Hjelsvold, R., & Midtstraum, R. (1994). Modelling and querying video data. *Proceedings of the 20th VLDB Conference*, Santiago, Chile (pp. 686-694).

ISO/IEC JTC1/SC29/WG11 (2001). Overview of the MPEG-7 Standard (version 5). Singapore, March.

Jain, R., & Hampapur, A. (1994). Metadata in video databases. *Sigmod Record*, *23*(4), 27-33.

Lindley, C., & Srinivasan, U. (1998). Query semantics for content-based retrieval of video data:

An empirical investigation. *Storage and Retrieval Issues in Image- and Multimedia Databases*, in conjunction with the *Ninth International Conference DEXA98,* August 24-28, Vienna, Austria.

Meng, J., Juan, Y., & Chang, S.-F. (1995). Scene change detection in a mpeg compressed video sequence. *SPIE Symposium on Electronic Imaging: Science & Technology- Digital Video Compression: Algorithms and Technologies*, 2419, San Jose, California, February.

Metz, C. (1974). *Film language: A semiotics of the cinema* (trans. by M. Taylor). The University of Chicago Press.

MPEG MAAATE (2000). The Australian MPEG Audio Analysis Tool Kit. Online at *http://www.cmis.csiro.au/dmis/maaate/*

Nepal, S., Srinivasan, U., & Reynolds, G. (2001). Automatic detection of "goal" segments in basketball videos. *ACM Multimedia 2001,* 261-269, Sept-Oct.

Nepal, S., Srinivasan, U., & Reynolds, G. (2001). Semantic-based retrieval model for digital audio and video. *IEEE International Conference on Multimedia and Exposition (ICME 2001),* August (pp. 301-304).

Pradhan, S., Tajima, K., & Tanaka, K. (2001). A query model to synthesize answer intervals from indexed video unit. *IEEE Transaction On Knowledge and Data Engineering, 13*(5), 824-838.

Rui, Y., Huang, T.S., & Chang, S.F (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation, 10,* 39-62.

Schloss, G.A., & Wynblatt, M.J.(1994). Building temporal structures in a layered multimedia data model. *ACM Multimedia*, 271-278.

Smith, J.R., & Benitez, A.B. (2000). Conceptual modeling of audio-visual content. *IEEE International Conference on Multimedia and Expo (ICME 2000),* July-Aug. (p. 915).

Smith, J.R., & Chang, S.-F. (1997). Querying by color regions using the VisualSEEk content-based visual query system. In M. T. Maybury (Ed.), *Intelligent multimedia information retrieval.* IJCAI.

Srinivasan, U., Gu, L., Tsui, & Simpsom-Young, B. (1997). A data model to support content-based search on digital video libraries. *The Australian Computer Journal, 29*(4), 141-147.

Srinivasan, U., Lindley, C., & Simpsom-Young, B. (1999). A multi-model framework for video information systems. *Database Semantics- Semantic Issues in Multimedia Systems,* January (pp. 85-108). Kluwer Academic Publishers.

Srinivasan, U., Nepal, S., & Reynolds, G. (2001). Modelling high level semantics for video data management. *Proceedings of ISIMP 2001,* Hong Kong, May (pp. 291-295).

Tansley, R., Dobie, M., Lewis, P., & Hall, W. (1999). MAVIS 2: An architecture for content and concept based multimedia information exploration. *ACM Multimedia*, 203.

Truong, B.T., Dorai, C., & Venkatesh, S. (2001). Determining dramatic intensification via flashing lights in movies. *International Conference on Multimedia and Expo,* August 22-25, Tokyo (pp. 61-64).

Yap, K., Simpson-Young, B., & Srinivasan, U. (1996). Enhancing video navigation with existing alternate representations. *First International Conference on Image Databases and Multimedia Search*, Amsterdam, August.

This work was previously published in Managing Multimedia Semantics, edited by U. Srinivasan and S. Nepal, pp. 135-159, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 7.7
Perceptual Multimedia:
A Cognitive Style Perspective

Gheorghita Ghinea
Brunel University, UK

Sherry Y. Chen
Brunel University, UK

ABSTRACT

In this chapter, we describe the results of empirical studies which examined the effect of cognitive style on the perceived quality of distributed multimedia. We use two dimensions of Cognitive Style Analysis, Field Dependent/Independent and Verbaliser/Visualiser, and the Quality of Perception metric to characterise the human perceptual experience. This is a metric which takes into account multimedia's infotainment (combined informational and entertainment) nature, and comprises not only a human's subjective level of enjoyment with regards to multimedia content quality, but also his/her ability to analyse, synthesise and assimilate the informational content of such presentations. Results show that multimedia content and dynamism are strong factors influencing perceptual quality.

INTRODUCTION

Multimedia has been identified as a potential method of improving the learning process in particular and the user computing experience in general. Its use encourages user interaction, thus ensuring that users cannot become a passive participant of the learning experience (Neo & Neo, 2004). Not only does the use of multimedia in applications increase interaction levels, but increases interest, motivation, and retention of information (Demetriadis, Triantafilou, & Pombortis, 2003). Moreover, the fact that students seem to prefer the use of multimedia for teaching to the standard teacher-student paradigm as it is more user-centred (Zwyno, 2003), comes as no surprise. The potential of multimedia has, to date, not been fully realised. Users perceive, process, and organise information in individualistic ways, yet current multimedia applications routinely fail to take this into consideration (Stash, Cristea, & De Bra, 2004).

We therefore believe that the effectiveness of a multimedia presentation would be hindered if it did not include the user experience in terms of enjoyment and information assimilation. Key to this is the issue of the quality of the multimedia presentation. Quality, in our perspective, has two main facets in a distributed multimedia environ-

ment: of service and of perception. The former, Quality of Service (QoS), illustrates the technical side of computer networking and represents the performance properties that the underlying network is able to provide. The latter, Quality of Perception (QoP), characterises the perceptual experience of the user when interacting with multimedia applications. While the quality delivered by communication networks has traditionally been measured using QoS metrics, we believe that, as users are "consumers" of multimedia applications, it is their opinions about the quality of multimedia material visualised which ultimately measures the success (or indeed, failure) of such applications to deliver desktop instruction material. When this delivery is done over Wide Area Networks such as the World Wide Web ("the Web"), transmission of multimedia data has to accommodate not only user subjective preferences, but also fluctuating networking environments.

The other concern is whether distributed multimedia presentations can accommodate individual differences. Previous studies indicate that users with different characteristics have different perceptions of multimedia presentation (Chen & Angelides, 2003). In particular, different cognitive style groups benefit from different types of multimedia presentation. Therefore, empirical evaluation that examines the impact of cognitive styles becomes paramount because such evaluations can provide concrete prescriptions for developing learner-centred systems that can match the particular needs of each cognitive style group. While QoP has been investigated in the context of distributed multimedia quality (Ghinea & Thomas, 2005), the study did not take into account the possible effect of users' cognitive styles on their QoP.

In this chapter, we present the results of studies which looked at how multimedia content is perceived by different cognitive style groups. Accordingly, the chapter begins by building a theoretical background to present previous work in the area of subjective distributed multimedia quality and to discuss the influence of cognitive style on user perception of multimedia presentations. It then describes and discusses the findings of our empirical studies. The chapter ends with conclusions being drawn, highlighting the value of integrating QoP considerations with users' cognitive styles in the delivery of distributed multimedia presentations.

THEORETICAL BACKGROUND

Quality of Service

The networking foundation on which current distributed multimedia applications are built either do not specify QoS parameters (also known as best effort service) or specify them in terms of traffic engineering parameters such as delay, jitter, and loss or error rates. However, these parameters do not convey application-specific needs such as the influence of clip content and informational load on the user multimedia experience.

Furthermore, traditional approaches of providing QoS to multimedia applications have focused on ways and means of ensuring and managing different technical parameters such as delay, jitter, and packet loss over unreliable networks. To a multimedia user, however, these parameters have little immediate meaning or impact. Although (s)he might be slightly annoyed at the lack of synchronisation between audio and video streams, it is highly unlikely that (s)he will notice, for instance, the loss of a video frame out of the 25 which could be transmitted during a second of footage, especially if the multimedia video in question is one in which the difference between successive frames is small. Moreover, in a distributed setting, the underlying communication system will not be able to provide an optimum QoS due to two competing factors, multimedia data sizes and network bandwidth. This results in phenomena such as congestion, packet loss, and errors. However, little work has

been reported on the relationship between the network provided QoS and the satisfaction and perception of the user.

While the QoS impacts upon the perceived multimedia quality in distributed systems, previous work examining the influence of varying QoS on user perceptions of quality has almost totally neglected multimedia's infotainment quality (i.e., a mixture of both of informational as well as entertainment content), and has concentrated primarily on the perceived entertainment value of presentations displayed with varying QoS parameters.

Accordingly, previous work has studied the impact of varying clip frame rates on the user's enjoyment of multimedia applications (Apteker, Fisher, Kisimov, & Neishlos, 1995; Fukuda, Wakamiya, Murata, & Miyahara, 1997), and it has been shown that the dependency between human satisfaction and the required bandwidth of multimedia clips is non-linear. Consequently, a small change in human receptivity leads to a much larger relative variation of the required bandwidth. From a different perspective, Wijesekera and Srivastava (1996) and Wijesekera, Srivastava, Nerode, and Foresti (1999) have examined the effect that random media losses have on the user-perceived quality. Their work showed that missing a few media units will not be negatively perceived by a user, as long as too many such units are not missed consecutively and that this occurrence is infrequent. Moreover, because of the bursty nature of human speech (i.e. talk periods interspersed with intervals of silence), audio loss is tolerated quite well by humans as it results merely in silence elimination (21% audio loss did not provoke user discontent (Wijesekera et al., 1999). However, viewer discontent for aggregate video losses increases gradually with the amount of losses, while for other types of losses and synchronisation defects, there is an initial sharp rise in viewer annoyance that afterwards plateaus out.

Further work has been undertaken by Steinmetz (1996) who explored the bounds within which lip synchronisation can fluctuate without undue annoyance on the viewer's part, while the establishment of metrics for subjective assessment of teleconferencing applications was explored in Watson and Sasse (1996). Indeed, the correlation between a user's subjective ratings of differing-quality multimedia presentation and physiological indicators has been studied by Wilson and Sasse (2000). However, research has largely ignored the influence that the user's psychological factors have on the perceived quality of distributed multimedia.

The focus of our research has been the enhancement of the traditional view of QoS with a user-level defined QoP. This is a measure that encompasses not only a user's satisfaction with multimedia clips, but also his/her ability to perceive, synthesise, and analyse the informational content of such presentations. As such, we have investigated the interaction between QoP and QoS and its implications from both a user perspective as well as from a networking angle.

Table 1. The differences between field-dependent and field-independent learners (Adapted from Jonassen & Grabowski, 1993; Riding & Rayner, 1998)

Field Dependent Learners	Field Independent Learners
They are externally directed and are easily influenced by salient features.	They are internally directed and process information with their own structure.
They experience surroundings in a relatively global fashion and struggle with individual elements.	They experience surroundings analytically and are good with problems that require taking elements out of their whole context.
They are more likely to accept ideas as presented.	They are more likely to accept ideas only strengthened through analysis.

Cognitive Styles

Cognitive style is an individual's characteristic and consistent approach to organising and processing information. Riding and Rayner (1998) defined cognitive style as "an individual preferred and habitual approach to organising and representing information" (p.25). Among a variety of cognitive styles, Field Dependence/Independence and Visualiser/Verbaliser are related to perceptual multimedia. The former concerns how users process and organise information, whereas the latter emphasises on how users perceive the presentation of information.

Field Dependence/Independence is related to the "degree to which a learner's perception or comprehension of information is affected by the surrounding perceptual or contextual field" (Jonassen & Grabowski, 1993, p. 87). Field-dependent people tend to perceive objects as a whole, whereas Field -independent people focus more on individual parts of the object. Field-dependent individuals rely more on external references; by contrast, field-independent individuals rely more on internal references (Witkin, Moore, Goodenough, & Cox, 1977). The differences between field-independent and field-dependent users are summarized in Table 1.

The main difference between visualiser/verbaliser focuses on a preference for learning with words versus pictures. A visualiser would prefer to receive information via graphics, pictures, and images, whereas a verbaliser would prefer to process information in the form of words, either written or spoken (Jonassen & Grabowski, 1993). In addition, visualisers prefer to process information by seeing, and they will learn most easily through visual and verbal presentations, rather than through an exclusively verbal medium. Moreover, their visual memory is much stronger than their verbal. On the other hand, verbalisers prefer to process information through words, and find they learn most easily by listening and talking (Laing, 2001). Their differences are summarised in Table 2.

These two dimensions of cognitive styles have been investigated by several works in learning environments. For example, a study by Chuang (1999) produced four-courseware versions: animation+text, animation+voice, animation+text+voice, and free choice. The result showed that field-independent subjects in the animation+text+voice group or in the free choice group scored significantly higher than those did in the animation+text group or in the animation+voice group. No significant presentation effect was found for the field-dependent subjects. Furthermore, Riding and Douglas (1993), with 15-16-year-old students, found that the computer-presentation of material on motorcar braking systems in a Text-plus-Picture format facilitated the learning by visualisers compared with the same content in a Text-plus-Text version. They further found that in the recall task in the Text-plus-Picture condition, 50% of the visualisers used illustrations as part of their answers, compared to only 12% of the verbalisers. Generally, visualisers learn best from pictorial presentations, while verbalisers learn best from verbal presentations. However, paucity of study investigates the relationship between the use patterns of these two dimensions of cognitive styles in multimedia systems in general, and specifically in distributed multimedia systems, where quality

Table 2. The differences between visualisers and verbalisers (Adapted from Jonassen & Grabowski, 1993; Riding & Rayner, 1998)

Visualisers	Verbalisers
Think concretely	Think abstractly
Have high imagery ability and vivid daydreams	Have low imagery ability
Like illustrations, diagrams, and charts	Like reading text or listening
Prefer to be shown how to do something	Prefer to read about how to do something
Are more subjective about what they are learning	Are more objective about what they are learning

fluctuations can occur owing to dynamically varying network conditions. As the QoP metric is one which has an integrated view of user-perceived multimedia quality in such distributed systems, it is of particular interest to investigate the impact of cognitive styles on QoS-mediated QoP, as it will help in achieving a better understanding of the factors involved in such environments (distance learning and CSCW, to name but two) and ultimately help in the elaboration of robust user models which could be used to develop applications that meet with individual needs.

METHODOLOGY DESIGN

Overview

This chapter reports the results of studies which investigated how different multimedia content is perceived by different cognitive style dimensions. Perceived multimedia quality was examined using the QoP measure, the only such metric that takes into account multimedia's infotainment duality.

Measuring QoP

As previously mentioned, QoP has two components: an information analysis, synthesis, and assimilation part (henceforth denoted by *QoP-IA*) and a subjective level of enjoyment (henceforth denoted by *QoP-LOE*). To understand QoP in the context of our work, it is important to explain how both these components were defined and measured.

Measuring Information Assimilation (QoP-IA)

In our approach, QoP-IA was expressed as a percentage measure, which reflected a user's level of information assimilated from visualised multimedia content. Thus, after watching a particular multimedia clip, the user was asked a standard number of questions (10, in our case) which examined information being conveyed in the clip just seen, and QoP-IA was calculated as being the proportion of correct answers that users gave to these questions. All such questions asked must, of course, have definite answers, for example: (from the Rugby video clip used in our experiments) "What teams are playing?" had an unambiguous answer (England and New Zealand) which had been presented in the multimedia clip, and it was therefore possible to determine if a participant had answered this correctly or not.

Thus, by calculating the percentage of correctly-absorbed information from different information sources, it was possible to determine from which information sources participants absorbed the most information. Using this data, it is possible to determine and compare, over a range of multimedia content, potential differences that might exist in QoP-IA.

Measuring Subjective Level of Enjoyment (QoP-LOE)

The subjective level of enjoyment (QoP-LOE) experienced by a user when watching a multimedia presentation, was polled by asking users to express, on a scale of 1-6, how much they enjoyed the presentation (with scores of 1 and 6 respectively representing "no" and "absolute" user satisfaction with the multimedia video presentation).

In keeping with the methodology followed by Apteker et al (1995), users were instructed not to let personal bias towards the subject matter in the clip or production-related preferences (for instance, the way in which movie cuts had been made) influence their enjoyment quality rating of a clip. Instead, they were asked to judge a clip's enjoyment quality by the degree to which they, the users, felt that they would be satisfied with a general purpose multimedia service of such quality. Users were told that factors which should influence their quality rating of a clip included clarity and acceptability of audio sig-

nals, lip synchronisation during speech, and the general relationship between visual and auditory message components. This information was also subsequently used to determine whether ability to assimilate information has any relation to user level of enjoyment, the second essential constituent (beside information analysis, synthesis, and assimilation) of QoP.

Participants

This chapter brings together the results of two empirical studies conducted at Brunel University's School of Information Systems, Computing, and Mathematics. In the first study, participants' cognitive styles were categorised according to the Field Dependent/Independent dimension, while in the second, participants' cognitive styles were categorised using the Verbaliser/Visualiser dimension.

Experiment 1

Sixty-six subjects participated in this study. Despite the fact that the participants volunteered to take part in the experiment, they were extremely evenly distributed in terms of cognitive styles, including 22 field-independent users, 22 intermediate users, and 22 field-dependent users. In terms of gender, there were 34 male users and 32 female users.

Experiment 2

This study involved 71 participants, which turned out to be quite evenly distributed in terms of cognitive styles, including 23 field-independent users, 25 intermediate users, and 23 field- dependent users. Moreover, participant breakdown according to gender was also quite evenly matched (37 males and 34 females). For both studies, all participating

Table 3. Video categories used in experiments

VIDEO CATEGORY	Dynamic	Audio	Video	Text
1 - Action Movie	Strong	Medium	Strong	Weak/None
2 - Animated Clip	Medium	Medium	Strong	Weak/None
3 - Band Clip	Medium	Strong	Medium	Weak/None
4 - Chorus Clip	Weak	Strong	Medium	Weak/None
5 - Commercial/Ad Clip	Medium	Strong	Strong	Medium
6 - Cooking Clip	Weak	Strong	Strong	Weak/None
7 - Documentary Clip	Medium	Strong	Strong	Weak/None
8 - News Clip	Weak	Strong	Strong	Medium
9 - Pop Music Clip	Medium	Strong	Strong	Strong
10 - Rugby Clip	Strong	Medium	Strong	Medium
11 - Snooker Clip	Weak	Medium	Medium	Strong
12 - Weather Forecast Clip	Weak	Strong	Strong	Strong

Figure 1. Snapshots of clips used in experiments

users were inexperienced in the content domain of the multimedia video clips visualised as part of our experiments, which will be described next.

Research Instruments

Video Clips

A total of 12 video clips were used in our study. The multimedia clips were visualised under a Microsoft Internet Explorer browser with a Microsoft Media player plug-in, with users subsequently filling in a Web-based questionnaire to evaluate QoP for each clip.

These 12 clips had been used in previous QoP experiments (Ghinea & Thomas, 1998), and were between 30-44 seconds long and digitised in MPEG-1 format. The subject matter they portrayed was varied (as detailed in Table 3 and Figure 1) and taken from selected television programmes, thereby reflecting informational and entertainment sources that average users might encounter in their everyday lives. Thus, six of the clips (2, 5, 6, 7, 8, and 12 in Table 1) comprised predominantly informational content, with the remainder of the clips being viewed mainly for entertainment purposes. Also varied was the dynamism of the clips (i.e., the rate of change between the frames of the clip), which ranged from a relatively static news clip to a highly dynamic space action movie. Table 3 also describes the importance, within the context of each clip, of the audio, video, and textual components as purveyors of information, as previously established through user tests (Ghinea & Thomas, 1998).

Cognitive Style Analysis

The cognitive style dimensions investigated in this study include Field Dependence/Independence and Verbaliser/Visualiser. A number of instruments have been developed to measure these two dimensions. Riding's (1991) Cognitive Style Analysis (CSA) was applied to identify each participant's cognitive styles in this study, because the CSA offers computerised administration and scoring. In addition, the CSA can offer various English versions, including Australasian, North American, and United Kingdom contexts.

The CSA uses two sub-tests to identify Field Dependence/Independence. The first presents items containing pairs of complex geometrical figures that the individual is required to judge as either the same or different. The second presents items each comprising a simple geometrical shape, such as a square or a triangle, and a complex geometrical figure, as in the GEFT, and the individual is asked to indicate whether or not the simple shape is contained in a complex one by pressing one of two marked response keys (Riding & Grimley, 1999).. The first sub-test is a task requiring field-dependent capacity. Conversely, the second sub-test requires the disembedding capacity associated with field-independence.

The CSA uses two types of statement to measure the Verbal-Imagery dimension and asks participants to judge whether the statements are true or false. The first type of statement contains information about conceptual categories while the second describes the appearance of items. There are 48 statements in total covering both types of statement. Each type of statement has an equal number of true statements and false statements. It is assumed that visualisers respond more quickly to the appearance statements, because the objects can be readily represented as mental pictures and the information for the comparison can be obtained directly and rapidly from these images. In the case of the conceptual category items, it is assumed that verbalisers have a shorter response time because the semantic conceptual category membership is verbally abstract in nature and cannot be represented in visual form. The computer records the response time to each statement and calculates the Verbal-Visualiser Ratio. A low ratio corresponds to a verbaliser and a high ratio to a visualiser, with the intermediate position being described as biomodal.

This study followed Riding's recommendation for the measurements of Field Dependence/Independence and Verbaliser/Visualiser. In terms of Field Dependence/Independence, Riding's (1991) recommendations are that scores below 1.03 denote field-dependent individuals; scores of 1.36 and above denote field-independent individuals; students scoring between 1.03 and 1.35 are classed as Intermediate. Regarding the measurement of Verbaliser/Visualiser, the recommendations are scores below 0.98 denote verbalisers; scores of 1.09 and above denote visualisers; students scoring between 0.98 and 1.09 are classed as biomodal.

Procedure

The experiment consisted of several steps. Initially, the CSA was used to classify users' cognitive styles as Field Dependent /Intermediate/Field Independent (Experiment 1) or Verbaliser/Biomodal/Visualiser (Experiment 2). Subjects then viewed the 12 multimedia video clips. In order to counteract any order effects, the order in which clips were visualised was varied randomly for each participant. After the users had seen each clip once, the window was closed, and they had to answer a number of questions about the video clip they had just seen. The actual number of such questions depended on the video clip, and varied between 10 and 12. After the user had answered the set of questions pertaining to a particular video clip and the responses had been duly recorded, (s)he was asked to rate the enjoyment quality of the clip that had just been seen on a Likert scale of 1 - 6 (with scores of 1 and 6 representing the worst and, respectively, best perceived qualities possible). The user then went on and watched the next clip.

Users were instructed not to let personal bias towards the subject matter in the clip or production-related preferences (for instance, the way in which movie cuts had been made) influence their enjoyment quality rating of a clip. Instead, they were asked to judge a clip's enjoyment quality by the degree to which they, the users, felt that they would be satisfied with a general purpose multimedia service of such quality. Users were told that factors which should influence their quality rating of a clip included clarity and acceptability of audio signals, lip synchronisation during speech, and the general relationship between visual and auditory message components.

Data Analyses

In this study, the independent variables include the participants' cognitive styles, as well as clip categories and their degree of dynamism. The dependent variables were the two components of Quality of Perception: the level of understanding (QoP-IA, expressed as a percentage measure describing the proportion of questions that the user had correctly answered for each clip) as well as the level of enjoyment (QoP-LOE, expressed

on a six-point Likert scale). Data were analysed with the Statistical Package for the Social Sciences (SPSS) for Windows version (release 9.0). An ANalysis Of VAriance (ANOVA), suitable to test the significant differences of three or more categories, and t-test, suitable to identify the differences between two categories (Stephen & Hornby, 1997), were applied to analyse the participants' responses. A significance level of $p < 0.05$ was adopted for the study.

DISCUSSION OF RESULTS

Subject Content

Experiment 1

Our analysis has highlighted that the subject content (i.e., particular clip category) has a statistically significant impact on the QoP-IA level of participants (p=.0000). This confirms previous results (Ghinea & Thomas, 1998) and we extended our analysis to include the impact of users' cognitive styles. As depicted in Table 4, the clip on which participants performed best varied according to the users' cognitive style. Accordingly, intermediate and field-independent users had the highest level of understanding for the Snooker video clip. However, field- dependent users perform better in the Documentary clip. As Table 1 shows, the most discriminating feature between the two clips is the fact that the latter clip does not contain any textual description. The difference in the understanding corresponding to the different cognitive styles is thus probably due to the fact that field-dependent learners are "influenced by salient features" (Jonassen & Grabowski, 1993, p.88). Without the distraction of text description, field-dependent users could concentrate their learning on video clips, so they could have better performance. On the other hand, irrespective of the type of cognitive style, all users performed worst in the highly dynamic Rugby sports action clip, in which, as Table 4 shows, all media components were medium-strong purveyors of information. This finding seems to imply that users have difficulty concentrating on multiple, different sources of information, and that the dynamism of the clip is also a contributing factor to participants' level of understanding, as was confirmed by further analysis, presented in the section titled "Degree of Dynamism"..

The specific multimedia clip type also influences the QoP-LOE, namely the level of enjoyment experienced by users (p=.0000). As Table 5 shows, although the Documentary and Rugby video clips predominate in the "Most Enjoyed" and "Least Enjoyed" categories, only for field-dependent users does the choice of most/least enjoyed clip coincide with the clips on which the level of understanding is highest, respectively lowest (see Table 4, in this regard). These results are in line with those of Fullerton (2000) and Ford and Chen (2001), which showed that field- dependent users performed better in a learning environment matching their preferences; conversely, their performance will be reduced in a mismatched condition. Moreover, our results also show that the Forecast video clip was the one which field-

Table 4. Cognitive styles and QoP-IA in clip categories

	Field Dependent	Intermediate	Field Independent
Best Performance	Documentary	Snooker	
	63.84%	62.93%	63.46%
Worst Performance		Rugby	
	33.24%	37.5%	34.82%

Table 5. Cognitive styles and QoP-LOE of clip categories

	Field Dependent	Intermediate	Field Independent
Most Enjoyed	Documentary		Forecast
	3.18	2.86	3.02
Least Enjoyed	Rugby	Band	Rugby
	1.39	1.93	1.61

independent users enjoyed most. It is probable that the wealth of detail present in this clip (information was conveyed through all three channels, video, audio, and text) is what makes the clip appealing to this particular type of users, who concentrated primarily on procedural details when processing information in a learning context (Pask, 1976, 1979). This is in contrast to the "Most Enjoyed" clip (i.e. Documentary) for the other two categories of users, in which information was only conveyed through the video and audio streams.

Experiment 2

Our results indicate that clip categories, as given by their specific multimedia content matter, significantly influence participants' components of QoP-IA. This shows that the information assimilation scores are significantly influenced by the content being visualised; moreover, this observation is valid irrespective of the particular cognitive style of the participant. However, closer analysis reveals that different cognitive style groups have different favourite clips. Pop Music, which displays information using multiple channels, including video, audio, and text, is the favourite clip, from an information assimilation point of view, for biomodals who combine the characteristics of both verbalisers and visualisers and are particularly adept at receiving information from either textual descriptions or graphic presentations. However, we did obtain some significant results that contradict those of previous research (Laing, 2001; Riding & Watts, 1997)

Table 6. Favourite clips

	Verbaliser	Biomodal	Visualiser
Score	62.84%	62.98%	65.52%
	Documentary	Snooker	
Enjoyment	4.17	3.94	3.91
	Documentary	Pop Music	Documentary

Table 7. Least favourite clip

	Verbaliser	**Biomodal**	**Visualiser**
Score	35.59%	35.74%	34.49%
		Rugby	
Enjoyment	2.59	2.78	2.81
		Rugby	

(Tables 6 and 7). Although the Documentary clip does not display any text description, it is the clip in which, on average, verbalisers obtain the highest QoP-IA (F=10.592, df=5,40, p=.000). On the other hand, visualisers perform better in the Snooker clip, which, though static, includes information conveyed through video, audio, and text (F=14.8451, df=6,36, p=.000).

However, irrespective of cognitive style, we found that the Rugby clip was the one in which participants obtained the lowest QoP-IA scores (F=32.743, df=15,72, p=.000). Although this clip is similar in some respects to others studied by us (such as the Snooker clip, which also has an abundance of information being portrayed through video, audio, and textual means), its main distinguishing feature is its high dynamism; there is considerable temporal variability due to the high inter-frame differences specific to clips featuring action sports. We therefore assume that the reason why participants scored so lowly in terms of QoP-IA on this clip is precisely because of its high dynamism, a hypothesis that shall be further explored in the section titled "Degree of Dynamism".

Enjoyment will also influence users' performance, especially for verbalisers, who perform better and enjoy more the Documentary clip and performed worse and enjoyed less the Rugby clip. It is consistent with the results of previous research (Chen, 2002), which highlight that positive user perceptions of the learning environment can enhance their performance; conversely, negative attitudes will tend to hinder learning performance.

Degree of Dynamism

Multimedia clip dynamism was found to significantly impact upon participants' QoP-IA levels in our study (Figure 2). The correct one is that the level of significance was found to be p=.000 for field-dependent users and p=.001 for intermediate and field-independent users. Clip dynamism is given by Table 3, where the terms strong, medium, and weak were coded with the values of 3, 2, and 1, respectively. All users performed worst in the clips with strong dynamism. In particular, field- dependent users do not perform as well as field-independent and intermediate users. As suggested by previous works (Chen, 2002; Chen & Macredie, 2002), field-dependent users' performance was hindered in situations where they need to extract cues by themselves. Thus, in multimedia clips with strong dynamism that provided too many cues, field-dependent users might find it difficult to select relevant cues.

The dynamism of the visualised clips also influenced the level of enjoyment experienced by participants (p=.000). If a per-cognitive style analysis is pursued, we find that the level of enjoyment is influenced by the dynamism of the multimedia clip for both field-independent (p=.004) and field-dependent (p=.000) users. As

Figure 2. Impact of dynamism and cognitive styles on participants' QoP-IA

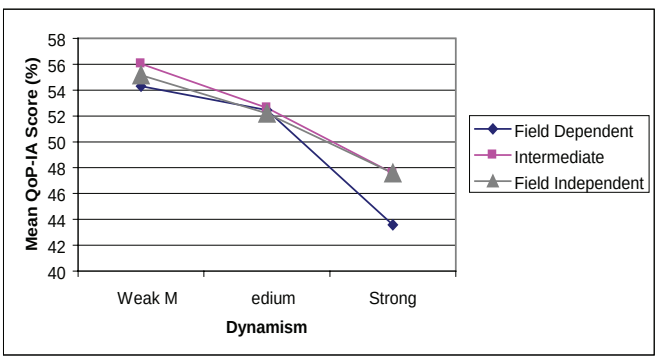
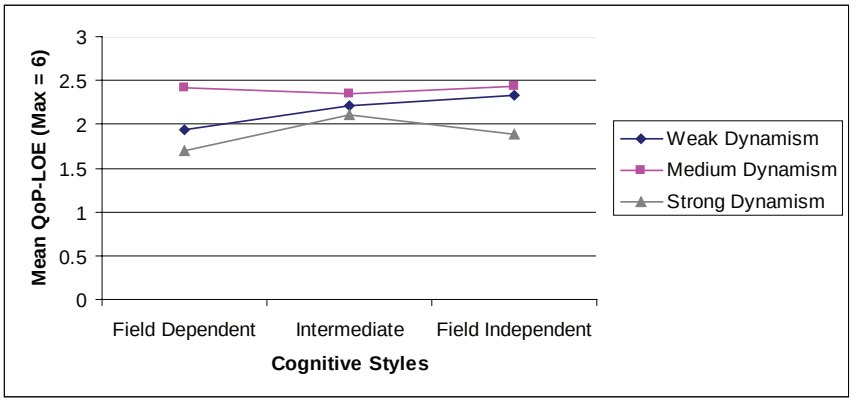


Figure 3. Impact of dynamism and cognitive styles on participants' QoP-LOE

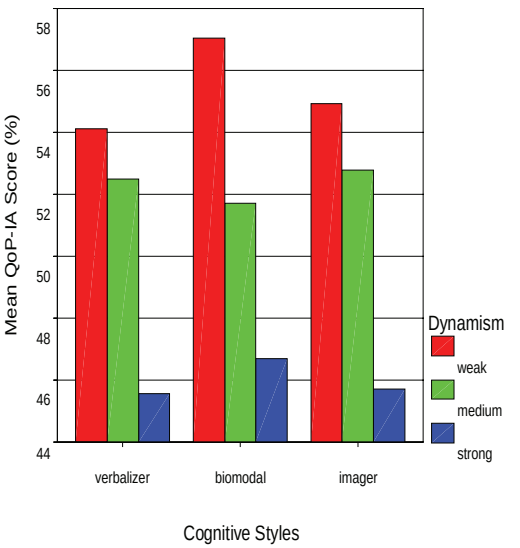


shown in Figure 3, both field-independent and field-dependent users experienced higher levels of enjoyment from the clips with medium dynamism, while strongly dynamic clips were liked least of all. However, dynamism does not seem to be a factor influencing multimedia clip enjoyment of intermediate users. One possible interpretation is that individuals possessing an Intermediate cognitive style employ a more versatile repertoire of information-seeking strategies. Versatile users, who have acquired the skill to move back and forth between different information-seeking strategies, are more capable of adapting themselves to suit the subject content presented by the multimedia video clips. This finding is consistent with the views of previous work, namely that a versatile strategy can be better equipped for multimedia learning technology (Chen & Ford, 1998; Paterson, 1996).

Analysis of the results obtained from the experiments shows that the degree of clip dynamism significantly impacts upon the QoP-IA component of QoP, irrespective of the user's cognitive style. The analysis has highlighted, moreover, the fact that the highest QoP-IA scores are obtained for clips which have a low degree of dynamism. Conversely, multimedia clips which have a high degree of dynamism have a negative impact on the user assimilation of the informational content being conveyed by the respective clips (Figure 4). Thus, clips which have relatively small inter-frame variability will facilitate higher QoP-IA scores: An object which might appear for only 0.5 seconds in a highly dynamic clip is less easily remembered than is the case when it appears for one second in a clip which is less dynamic.

As far as the QoP-LOE component is concerned, analysis of our results reveals that while

Figure 4. Cognitive styles and clip dynamism impact on QoP-IA



dynamism is a significant factor in the case of verbalisers and visualisers, this is not true of biomodals (Figure 5). As suggested by previous research (Riding & Rayner, 1998), biomodals can tailor learning strategies to the specific learning environments so the features of learning environments have no significant effects on their enjoyment. For verbalisers and visualisers, however, it was found that clips of medium dynamism had the highest levels of QoP-LOE, which suggests that such users do not find enjoyable clips which are static (or, conversely, highly dynamic). While the user may feel somewhat overwhelmed by a fast-paced clip, (s) he might possibly feel uninterested by a static clip with almost repetitive frame displays; it should come as no surprise, then, that such users prefer clips of medium dynamism, where they do not feel overwhelmed, but neither are they bored by the presentation of the subject matter concerned.

SUMMARY

This chapter has presented the results of two studies that looked at the impact of cognitive styles and multimedia content on users' subjective Quality of Perception. The latter is a term which encompasses not only a user's enjoyment, but also his/her level of understanding of visualised multimedia content. Each of the studies used a different dimension for categorising participants' cognitive style – thus, the first study used the Field Dependent/ Independent categorisation, whilst the second employed the Verbaliser/Visualiser dimension.

Our results reveal that multimedia video clip dynamism is an important factor impacting, irrespective of the particular dimension of cognitive style being employed, upon participants' QoP-IA levels. A similar conclusion as regards QoP-LOE can only be made, however, if the Field Dependent/Independent dimension is used. If one uses the Verbaliser/Visualiser dimension to classify cognitive style, clip dynamism has no significant effects on Biomodals, which, displaying characteristics of both Verbalisers and Visualisers, have adaptable preferences of accessing information and enjoy receiving information from multiple channels.

However, what our results do highlight, independent of the cognitive style taxonomy being used, is that multimedia content does have a significant influence on the user experience, as measured by the QoP metric. This would imply that in order to deliver an enhanced multimedia infotainment experience, multimedia content providers should focus on relatively static multimedia video and take into consideration the appropriateness of the subject matter in order to aid in the uptake and proliferation of distributed multimedia.

REFERENCES

Apteker, R. T., Fisher, J. A., Kisimov, V. S., & Neishlos, H. (1995). Video acceptability and frame rate. *IEEE Multimedia*, *2*(3), 32-40.

Chen, S. Y. (2002). A cognitive model for non-linear learning in hypermedia programmes. *British Journal of Educational Technology, 33*(4), 453-464.

Chen, S. Y., & Angelides, M. C. (2003). Customisation of Internet multimedia information systems design through user modelling. In S. Nansi (Ed.), *Architectural issues of Web-enabled electronic business* (pp. 241-255). Hershey, PA: Idea Group Publishing.

Chen, S. Y., & Ford, N. J. (1998). Modelling user navigation behaviours in a hypermedia-based learning system: An individual differences approach. *International Journal of Knowledge Organization, 25(3),* 67-78.

Chen, S. Y., & Macredie, R. D. (2002). Cognitive styles and hypermedia navigation: Development

of a learning model. *Journal of the American Society for Information Science and Technology, 53(1),* 3-15.

Chuang, Y. R. (1999). *Teaching in a multimedia computer environment: A study of effects of learning style, gender, and math achievement.* Retrieved December 21, 2001, from http://imej.wfu.edu./articles/1999/1/10/index/asp

Demetriadis, S., Triantafilou, E., & Pombortis, A. (2003, June 30-July 2). A phenomenographic study of students' attitudes toward the use of multiple media for learning. *ACM SIGCSE Bulletin, Proceedings of the 8th Annual Conference on Innovation and Technology in Computer Science Education,* Thessaloniki, Greece (pp. 183-187).

Ford, N., & Chen, S. Y. (2001). Matching/mismatching revisited: An empirical study of learning and teaching styles. *British Journal of Educational Technology, 32(1),* 5-22.

Fukuda, K., Wakamiya, N., Murata, M., & Miyahara, H. (1997). QoS mapping between user's preference and bandwidth control for video transport. *Proceedings of the 5th International Workshop on QoS (IWQoS)* (pp. 291-301).

Fullerton, K. (2000). *The interactive effects of field dependence-independence and Internet document manipulation style on student achievement from computer-based instruction.* Doctoral dissertation, University of Pittsburgh.

Ghinea, G., & Thomas, J. P. (2005). Quality of perception: User quality of service in multimedia presentations. *IEEE Transactions on Multimedia, 7(4),* 786-789.

Jonassen, D. H., & Grabowski, B. (1993). *Individual differences and instruction.* New York: Allen & Bacon.

Laing, M. (2001). Teaching learning and learning teaching: An introduction to learning styles. *New Frontiers in Education,* 31(4), 463-475.

Neo, T., & Neo, M. (2004). Classroom innovation: Engaging students in interactive multimedia learning. *Campus-Wide Information Systems, 21(3),* 118-124.

Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology, 46,* 128-148.

Pask, G. (1979). *Final report of S.S.R.C. Research Programme HR 2708.* Richmond (Surrey): System Research Ltd.

Paterson, P. (1996). *The influence of learning strategy in a computer mediated learning environment.* Paper presented at ALT-Conference '96. Retrieved November 19, 1997, from http://www.warwick.ac.uk/alt-/alt-96/papers.html

Riding, R. J. (1991). *Cognitive styles analysis.* Birmingham: Learning and Training Technology.

Riding, R. J., & Douglas, G. (1993). The effect of cognitive style and mode of presentation on learning performance. *British Journal of Educational Psychology, 63,* 297-307.

Riding, R. J., & Grimley, M. (1999). Cognitive style and learning from multimedia materials in 11 year old children. *British Journal of Educational Technology, 30(2),* 43-56.

Riding, R. J., & Rayner, S. G. (1998). *Cognitive styles and learning strategies.* London: David Fulton Publisher.

Riding, R. J., & Watts, M. (1997). The effect of cognitive style on the preferred format of instructional material. *Educational Psychology, 17*(1 & 2), 179-183.

Stash, N., Cristea, A., & De Bra, P. (2004, May). Authoring of learning styles in adaptive hypermedia: Problems and solutions. *Proceedings of the Thirteenth International World Wide Web Conference,* New York, (pp 114-123).

Steinmetz, R. (1996). Human perception of jitter and media synchronisation. *IEEE Journal on Selected Areas in Communications, 14*(1), 61-72.

Watson, A., & Sasse, M. A. (1996). Evaluating audio and video quality in low cost multimedia conferencing systems. *Interacting with Computers, 8*(3), 255-275.

Wijesekera, D., & Srivastava, J. (1996). Quality of service (QoS) metrics for continuous media. *Multimedia Tools and Applications, 3*(1), 127-136.

Wijesekera, D., Srivastava, J., Nerode, A., & Foresti, M. (1999). Experimental evaluation of loss perception in continuous media. *Multimedia Systems, 7*(6), 486-499.

Wilson, G. M., & Sasse, M. A. (2000, December). Investigating the impact of audio degradations on users: Subjective vs. objective assessment methods. *Proceedings of OZCHI2000*, Sydney (pp. 135-142).

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox., P. W. (1977). Field-dependent and field independent cognitive styles and their educational implications. *Review of Educational Research, 47*, 1-64.

Zwyno, M. S. (2003, November). Student learning styles, Web use patterns, and attitudes towards hypermedia-enhanced instruction. *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference* (pp. 1-6).

This work was previously published in Digital Multimedia Perception and Design, edited by G. Ghinea, and G. Ghinea, pp. 187-205, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.8
Incorporating User Perception in Adaptive Video Streaming Systems

Nicola Cranley
University College Dublin, Ireland

Liam Murphy
University College Dublin, Ireland

ABSTRACT

There is an increasing demand for streaming video applications over both the fixed Internet and wireless IP networks. The fluctuating bandwidth and time-varying delays of best-effort networks makes providing good quality streaming a challenge. Many adaptive video delivery mechanisms have been proposed over recent years; however, most do not explicitly consider user-perceived quality when making adaptations, nor do they define what quality is. This chapter describes research that proposes that an optimal adaptation trajectory through the set of possible encodings exists, and indicates how to adapt transmission in response to changes in network conditions in order to maximize user-perceived quality.

INTRODUCTION

Best-effort IP networks are unreliable and unpredictable, particularly in a wireless environment. There can be many factors that affect the quality of a transmission, such as delay, jitter, and loss. Congested network conditions result in lost video packets, which, as a consequence, produce poor quality video. Further, there are strict delay constraints imposed by streamed multimedia traffic. If a video packet does not arrive before its playout time, the packet is effectively lost. Packet losses have a particularly devastating effect on the smooth continuous playout of a video sequence due to inter-frame dependencies. A slightly degraded quality but uncorrupted video stream is less irritating to the user than a randomly-corrupted stream. However, rapidly fluctuating quality should also be avoided as the

human vision system adapts to a specific quality after a few seconds, and it becomes annoying if the viewer has to adjust to a varying quality over short time scales (Ghinea, Thomas, & Fish, 1999). Controlled video quality adaptation is needed to reduce the negative effects of congestion on the stream while providing the highest possible level of service and quality. For example, consider a user watching some video clip; when the network is congested, the video server must reduce the transmitted bitrate to overcome the negative effects of congestion. In order to reduce the bitrate of the video stream, the quality of the video stream must be reduced by sacrificing some aspect of the video quality. There are a number of ways in which the quality can be adapted; for example, the image resolution (i.e. the amount of detail in the video image), the frame rate (i.e. the continuity of motion), or a combination of both can be adapted. The choice of which aspect of the video quality should depend on how the quality reduction will be perceived.

In the past few years, there has been much work on *video quality adaptation* and *video quality evaluation*. In general, video quality adaptation indicates how the bit rate of the video should be adjusted in response to changing network conditions. However, this is not addressed in terms of video quality, as for a given bit rate budget there are many ways in which the video quality can be adapted. Video quality evaluation measures the quality of video as perceived by the users, but current evaluation approaches are not designed for adaptive video streaming transmissions.

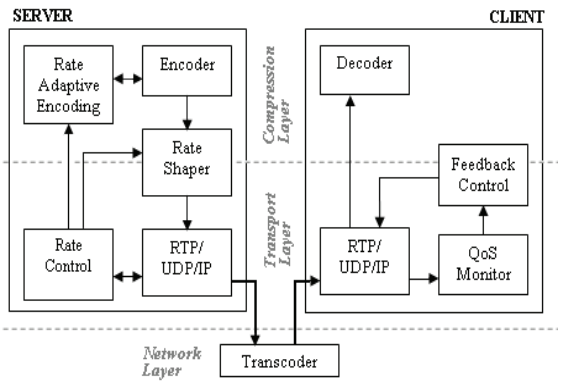
This chapter will firstly provide a generalized overview of adaptive multimedia systems and describe recent systems that use end-user perception as part of the adaptation process. Many of these adaptive systems rely on objective metrics to calculate the user-perceived quality. Several objective metrics of video quality have been developed, but they are limited and not satisfactory in quantifying human perception. Further, it can be argued that to date, objective metrics were not designed to assess the quality of an adapting video stream. As a case study, the discussion will focus on recent research that demonstrates how user-perceived quality can be used as part of the adaptation process for multimedia. In this work, the concept of an Optimal Adaptation Trajectory (OAT) has been proposed. The OAT indicates how to adapt multimedia in response to changes in network conditions to maximize user-perceived quality. Finally experimental subjective testing results are presented that demonstrate the dynamic nature of user-perception with adapting multimedia. The results illustrate that using a two-dimensional adaptation strategy based on the OAT out-performs one-dimensional adaptation schemes, giving better short-term and long-term user-perceived quality.

REVIEW OF ADAPTIVE MULTIMEDIA SYSTEMS

Given the seriousness of congestion on the smooth continuous play-out of multimedia, there is a strong need for adaptation. The primary goals of adapting multimedia are to ensure graceful quality adaptation, maintain a smooth continuous play-out and maximize the user-perceived quality. Multimedia servers should be able to intelligently adapt the video quality to match the available resources in the network. There are a number of key features that need to be considered in the development of an adaptive streaming system (Wang & Schulzrinne, 1999) such as feedback to relay the state of the network between client and server, the frequency of this feedback, the adaptation algorithm used, the sensitivity of the algorithm to feedback, and the resulting user-perceived quality.

However, the most important thing is how the system reacts, how it adapts to congestion, and the perceived quality that results from this adaptation.

Figure 1. Adaptation techniques



Adaptation Techniques

Broadly speaking, adaptation techniques attempt to reduce network congestion by matching the rate of the multimedia stream to the available network bandwidth. Without some sort of rate control, any data transmitted exceeding the available bandwidth would be discarded, lost, or corrupted in the network. Adaptation techniques can be classified into the following generalized categories: rate control, rate shaping, and rate adaptive encoding (Figure 1). Each of these techniques adapts the transmitted video stream to match the available resources in the network by either adapting the rate at which packets are sent or adjusting the quality of the delivered video (Wu, Hou, Zhu, Lee, Chiang, Zhang, & Chao, 2000, 2002). These are briefly described in the following sections.

Rate Control

Rate control is the most commonly-used mechanism employed in adaptive multimedia systems. Rate control can be implemented either at the server, the client, or a hybrid scheme whereby the client and server cooperate to achieve rate control.

- **Sender-based rate control:** On receipt of feedback from the client, the server adapts the transmission rate of the multimedia stream being transmitted in order to minimize the levels of packet loss at the client by matching the transmission rate of the multimedia stream to the available network bandwidth. Without any rate control, the data transmitted exceeding the available bandwidth would be discarded in the network.
- **Receiver-based rate control:** The clients control the receiving rate of video streams by adding/dropping layers. In layered multicast, the video sequence is compressed into multiple layers: a base layer and one or more enhancement layers. The base layer can be independently decoded and provides basic video quality; the enhancement layers can only be decoded together with the base layer, and they enhance the quality of the base layer.
- **Hybrid rate control:** This consists of rate control at both the sender and receiver. The hybrid rate control is targeted at multicast video and is applicable to both layered video and non-layered video. Typically, clients regulate the receiving rate of video streams by adding or dropping layers while the sender also adjusts the transmission rate of each layer based on feedback information from the receivers.

Unlike server-based schemes, the server uses multiple layers, and the rate of each layer may vary due to the hybrid approach of adapting both at the server and receiver.

Rate Shaping

Rate shaping is a technique to adapt the rate of compressed video bit-streams to meet some target bit rate by acting as a filter (or interface) between the compression layer and the transport layer. There are a number of filters that can be used to achieve rate shaping.

- **Frame-dropping filter:** This filter distinguishes between the different frame types in a video stream (i.e., I-, P- and B-frames). The frame-dropping filter is used to reduce the data rate of a video stream by discarding frames according to their relative importance. For example, B-frames are preferentially dropped, followed by P-frames and finally I-frames.
- **Frequency filter:** This filter performs filtering operations on the compression layer, for example, by discarding DCT coefficients at higher frequencies or reducing the color depth.
- **Re-quantization filter:** Re-quantizes the DCT coefficients. The filter extracts and de-quantizes the DCT coefficients from the compressed video stream then re-quantizes the coefficients with a larger quantization step which results in a reduced bitrate and reduced quality.

Rate Adaptive Encoding

Rate adaptive encoding performs adaptation by adjusting the encoding parameters which in turn adapts the output bit rate. However, adaptive encoding is constrained by the capabilities of the encoder and the compression scheme used. There are a number of encoding parameters that can be adapted in rate adaptive encoding, such as dynamically adapting the quantization parameter, frame rate, and/or the spatial resolution.

Discussion

The key questions that arise when developing or designing adaptation algorithms are how the system adapts and the perceived quality at the receiver.

There are a number of common components in each of the different adaptation techniques described. Many adaptation algorithms have a strong dependency on the choice of control parameters used within the adaptation process. For example, in a server-based rate control system, upon receipt of feedback the server either increases its transmission rate by α or decreases its rate by β. If the rate of α is chosen to be too large, the increased transmission rate could push the system into causing congestion, which can in turn cause the client to experience loss and poor perceived quality. However, if α is too small, the server will be very slow to make use of the extra available bandwidth and send a higher bit rate video stream. Thus, the algorithm is heavily dependent on the value of the control parameters, α and β, which drive the adaptation.

Even more problematic is translating rate into real video encoding parameters. Consider a simple system where the server is delivering video at 150kbps, and based on feedback, the algorithm indicates that the transmission rate should be increased to 160kps. The question that remains is: How should the extra 10kps be achieved, how can the video stream be adjusted to achieve this rate? This is further complicated by the limitations of the encoder to adapt the video. Layer-based schemes are equally problematic since there is no firm definition of what constitutes a base layer and each of the enhancement layers.

The most important issue that is often omitted in the design of adaptation algorithms is user-perception. User-perception should be incorporated into the adaptation algorithms, since it is the user who is the primary entity affected by adaptation, and should therefore be given priority in the adaptation decision-making process. For example, if a video clip is being streamed at a particular encoding configuration and the system needs to degrade the quality being delivered, how this adaptation occurs should be dictated by the users' perception. The way to degrade should be such as to have the least negative impact on the users' perception. There needs to be some sort of understanding of video quality and the perception of the video quality in order for adaptation to occur in an achievable and intelligent manner.

REVIEW OF OBJECTIVE METRICS

The main goal of objective metrics is to measure the perceived quality of a given image or video. Sophisticated objective metrics incorporate perceptual quality measures by considering the properties of the ***Human Visual System*** (HVS) in order to determine the visibility of distortions and thus the perceived quality. However, given that there are many factors that affect how users perceive quality, such as video content, viewing distance, display size, resolution, brightness, contrast, sharpness/fidelity, and colour, many objective metrics have limited success in calculating the perceived quality accurately for a diverse range of testing conditions and content characteristics. Several objective metrics of video quality have been proposed (Hekstra, 2002; van den Branden Lambrecht, 1996; Watson, Hu, & McGowan, 2000; Winkler, 1999), but they are limited and not satisfactory in quantifying human perception (Masry & Hemami, 2002; Yu & Wu, 2000).

In this section two key objective metrics, the ***Peak Signal to Noise Ratio*** (PSNR) and the ***Video Quality Metric*** (VQM) are reviewed. These two metrics have been widely applied to many applications and adaptation algorithms to assess video quality.

Peak Signal to Noise Ratio (PSNR)

The most commonly-used objective metric of video quality assessment is the ***Peak Signal to Noise Ratio*** (PSNR). The advantage of PSNR is that it is very easy to compute. However, PSNR does not match well to the characteristics of HVS. The main problem with using PSNR values as a quality assessment method is that even though two images are different, the visibility of this difference is not considered. The PSNR metric does not take the visual masking phenomenon or any aspects of the HVS into consideration, that is, every single errored pixel contributes to the decrease of the PSNR, even if this error is not perceived. For example, consider an image where the pixel values have been altered slightly over the entire image and an image where there is a concentrated distortion in a small part of the image both will result in the PSNR value however, one will be more perceptible to the user than the other. It is accepted that the PSNR does not match well to the characteristics of the HVS (Girod, 1993; van den Branden Lambrecht & Verscheure, 1996).

Video Quality Metric (VQM)

The ITU-T has recently accepted the Video Quality Metric (VQM) from the National Telecommunications and Information Administration (NTIA) as a recommended objective video quality metric that correlates adequately to human perception in ITU-T J.148 (2003) and ITU-T J.149 (2004). The ***Video Quality Metric*** (VQM) provides a means of objectively evaluating video quality. The system compares an original video clip and a processed video clip and reports a Video Quality Metric (VQM) that correlates to the perception of a typical end user. The VQM objective metrics are claimed to provide close approximations to the overall quality impressions, or mean opinion scores (Wolf & Pinson, 1999). The quality measurement process includes sampling of the original and processed video streams, calibration of the original and processed video streams, extraction of perception-based features, computation of video quality parameters, and finally calculation using various VQM models.

Using Objective Metrics for Multimedia Adaptation

Given the wide range of video quality metrics developed, the ***Video Quality Experts Group*** (VQEG) was formed in 1997 with the task of collecting reliable subjective ratings for a defined set of test sequences and to evaluate the performance of various objective video quality metrics (VQEG,

2005). In 2000, the VQEG performed a major study of various objective metrics on behalf of the ITU to compare the performances of various objective metrics against subjective testing in terms of prediction accuracy, prediction monotonicity, and prediction consistency. The results of the VQEG study found that no objective metric is able to fully replace subjective testing, but even more surprisingly, that no objective metric performed statistically better than the PSNR metric.

The main difficultly with video quality metrics is that even though they give an indication of the video quality, they do not indicate how the video quality should be adapted in an adaptive system. Furthermore, many of these objective metrics require a comparison between the reference clip and the degraded video clip in order to calculate the video quality. This comparison is often done on a frame-by-frame basis and therefore requires both the reference and degraded clips to have the same frame rate. The more sophisticated metrics proposed are extremely computationally intense and are unsuitable for use in a real-time adaptive system. Given the limitations of objective metrics, it has been recognized that user-perception needs to be incorporated in adaptation algorithms for streamed multimedia. There are emerging adaptive streaming systems being developed that address this issue (Muntean, Perry, & Murphy, 2004; Wang, Chang, & Loui, 2004).

OPTIMUM ADAPTATION TRAJECTORIES (OATS)

This section will focus on an approach that incorporates user-perception into adaptation algorithms for video streaming. This work proposes that there is an optimal way in which multimedia transmissions should be adapted in response to network conditions to maximize the user-perceived quality (Cranley, Murphy, & Perry, 2003). This is based on the hypothesis that within the set of different ways to achieve a target bit rate, there exists an encoding configuration that maximizes the user-perceived quality. If a particular multimedia file has *n* independent encoding configurations, then there exists an adaptation space with *n* dimensions. When adapting the transmission from some point within that space to meet a new target bit rate, the adaptive server should select the encoding configuration that maximizes the user-perceived quality for that given bit rate. When the transmission is adjusted across its full range, the locus of these selected encoding configurations should yield an ***Optimum Adaptation Trajectory*** (OAT) within that adaptation space.

This approach is applicable to any type of multimedia content. The work presented here focuses for concreteness on the adaptation of MPEG-4 video streams within a finite two-dimensional adaptation space defined by the range of the chosen encoding configurations. Each encoding configuration consists of a combination of frame rate and resolution and is denoted as [Frame rate $_{FPS}$, Resolution $_{R}$]. These encoding variables were chosen as they most closely map to the spatial and temporal complexities of the video content. The example shown in Figure 2(a) indicates that, when degrading the quality from an encoding configuration of 25fps and 100% resolution or [25_{FPS}, 100_{R}], there are a number of possibilities such as reducing the frame rate only, [X_{FPS}, 100_{R}], reducing the resolution only, [25_{FPS}, Y_{R}], or reducing a combination of both parameters, [U_{FPS}, V_{R}]. Each of these possibilities lies within a zone of ***Equal Average Bit Rate*** (EABR). The clips falling within a particular zone of EABR have different, but similar bit rates. For example, the bit rates corresponding to the encoding points [17_{FPS}, 100_{R}], [25_{FPS}, 79_{R}] and [25_{FPS}, 63_{R}] were 85, 88, and 82 kbps, respectively. To compare clips of exactly the same bit rate would require a target bit rate to be specified, and then the encoder would use proprietary means to achieve this bit rate by compromising the quality of the encoding in an unknown manner. Using zones of EABR effectively quantizes the bit rate of dif-

Figure 2(a). Adaptation possibilities

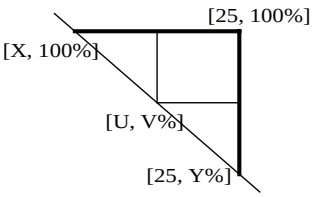
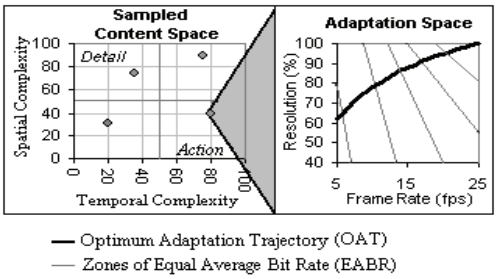


Figure 2(b). Adaptation space



ferent video sequences with different encoding configurations. The boundaries of these zones of EABR are represented as linear contours for simplicity, since their actual shape is irrelevant for this scheme.

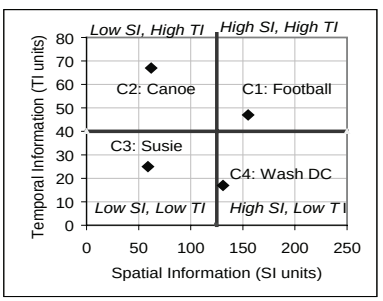
The OAT indicates how the quality should be adapted (upgraded or downgraded) so as to maximize the user-perceived quality. The OAT may be dependent on the characteristics of the content. There is a content space in which all types of video content exist in terms of ***spatial and temporal complexity*** (or *detail* and *action*). Every type of video content within this space can be expanded to an adaptation space as shown in Figure 2(b). Adaptation space consists of all possible dimensions of adaptation for the content. It can be implemented as part of an adaptive streaming server or adaptive encoder.

OAT Discovery

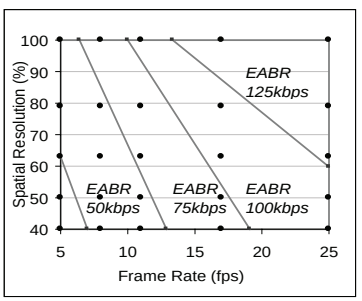
User perception of video quality may vary with the content type; for example, viewers may perceive action clips differently from slow-moving clips. Thus, there may exist a different OAT for different types of content based on their spatial and temporal characteristics. In order to characterize content in terms of its spatial and temporal complexity, a spatial-temporal grid was constructed, as shown in Figure 3(a). The spatial and temporal perceptual information of the content was determined using the metrics Spatial Information (SI) and Temporal Information (TI) (ITU-T P.910, 1999).

Eight different content types were selected based on their SI and TI values in order to cover as much of the Spatial-Temporal grid as possible. These test sequences were acquired from the VQEG. Each test sequence was then expanded to form an adaptation space, as shown in Figure

Figure 3(a). Spatial-temporal grid sampled with four content types for phase one of testing; (b) logarithmically-sampled adaptation space for content type C1

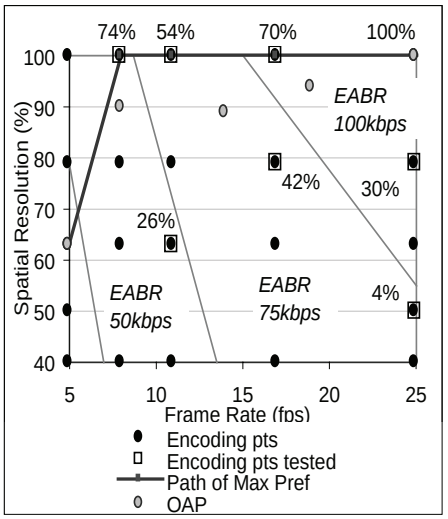


(a)



(b)

Figure 4. Subjective test results for content type, C3

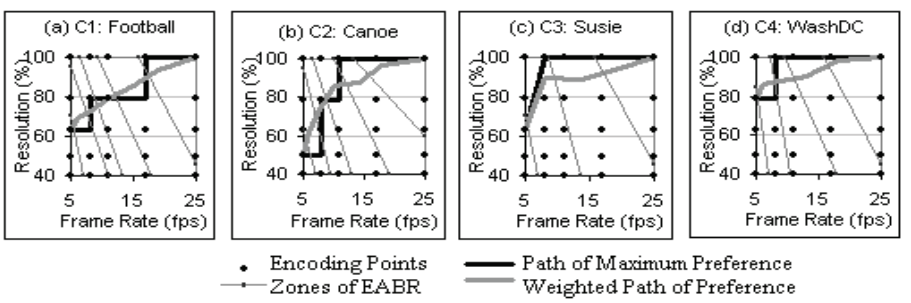


3(b). During the preparation of the test sequences for the subjective testing, the encoding method used was the "*most accurate*", that is, no target bit rate was specified, and the encoder followed the supplied encoding parameters as closely as possible regardless of the resulting bit rate.

The subjective testing consisted of two independent testers performing identical test procedures and using identical test sequences on subjects. Subjects were eliminated if the subject was either knowledgeable about video quality assessment or had any visual impairments. Testing was conducted in two phases. Phase One considered four test sequences, one taken from each quadrant of the SI-TI grid. To facilitate subjective testing and reduce the number of test cases, adaptation space was sampled using a logarithmic scale to reflect Weber's Law of Just Noticeable Difference (JND). Phase Two considered four different test sequences with similar SI-TI values to those used for Phase One. However, this time, the adaptation space was sampled using a linear scale. The main objective of having two different test phases was to verify and validate the results from Phase One. In addition, by using different encoding scales, it could be verified that the OAT was similar in shape regardless of whether a linear or logarithmic scale was used, and regardless of the encoding points tested.

There are a number of different subjective testing methodologies that are proposed by the ITU-T, including the ***Absolute Category Rating*** (ACR), the ***Degraded Category Rating*** (DCR), and ***Pair Comparison*** (PC) methods. The DCR method uses a five-point impairment scale whilst the ACR method uses a five-point quality grading scale, or alternatively a ***Continuous Quality Scale*** (CQS) (ITU-T P.910, 1999). However, by using such grading scales, it is criticized that different subjects may interpret the associated grading scale in different ways and use the grading scale in a non-uniform fashion (Watson, 1998). To overcome these difficulties in the grading procedure, the ***Forced Choice*** methodology is often employed. In the forced choice method, the subject is presented with a number of spatial or temporal alternatives in each trial. The subject is forced to choose the

Figure 5. Path of maximum user preference and weighted path of preference for four different content types



location or interval in which their preferred stimulus occurred. Using the forced choice method, the bias is binary, which simplifies the rating procedure and allows for reliability, verification, and validation of the results. The subjective tests consisted of a subject watching every combination of pairs of clips from each EABR zone for each content type and making a forced choice of the preferred encoding configuration. Intra-reliability and inter-reliability of a subject were factored into the test procedure by including repetition of the same test sequence presentation.

The diagram in Figure 4 shows the subjective test results obtained for a particular content type. The diagram consists of a grid of circular encoding points where the frame rate is on the x-axis and the resolution is on the y-axis. Through these encoding points are diagonal grey lines denoting the zones of EABR, ranging from 100kbps to 25kbps. The encoding points marked with a percentage preference value are those points that were tested within a zone of EABR. For example, in EABR-100kbps, there were two encoding configurations tested, [17_{FPS}, 100_{R}] and [25_{FPS}, 79_{R}]. Seventy percent of the subjects preferred encoding configuration [17_{FPS}, 100_{R}], while the remaining 30% preferred encoding configuration [25_{FPS}, 79_{R}]. However, in the left-most zone of EABR, the preferred encoding configuration is [5_{FPS}, 63_{R}]. In this zone of EABR there are three encoding configurations, but since the frame rate is the same, the preferred encoding configuration is that with the highest resolution, [5_{FPS}, 63_{R}].

The ***Path of Maximum Preference*** is the path through the zones of EABR joining the encoding configurations with the maximum user preference. Weighted points were then used to obtain the ***Optimal Adaptation Perception (OAP)*** points. The weighted points were interpolated as the sum of the product of preference with encoding configuration. For example, 70% of subjects preferred encoding [17_{FPS}, 100_{R}] and 30% preferred encoding point [25_{FPS}, 79_{R}]. The weighted vector of these two encoding configurations is [$70\%(17_{FPS})+30\%(25_{FPS})$, $70\%(100_{R})+30\%(79_{R})$] which equals OAP point [19.4_{FPS}, 93.7_{R}]. The ***Weighted Path of Preference*** is the path joining the OAPs. There are two possible paths which can be used to represent the OAT: the path of maximum user preference, and the weighted path of preference. It seems likely that by using the weighted path of preference, the system can satisfy more users by providing a smooth graceful quality adaptation trajectory. Using the same subjective testing methodology, the OAPs in each zone of EABR were compared against the maximum preferred encoding and all other encoding configurations. In all cases, the interpolated OAP did not have a statistically-significant preference from the maximum preferred encoding indicating that this simple weighted vector approach is acceptable. It was also observed that there was a higher incidence of forced choices when the maximum preferred encoding and the OAP were close together.

Figure 5 shows the paths of maximum preference and weighted paths of preference for the four content types used during Phase One of testing. It can be clearly seen from the paths of maximum user preference that when there is high action (C1 and C2), the resolution is less dominant regardless of whether the clip has high spatial characteristics or not. This implies that the user is more sensitive to continuous motion when there is high temporal information in the video content. Intuitively this makes sense as when there is high action in a scene; often the scene changes are too fast for the user to be able to assimilate the scene detail. Conversely, when the scene has low temporal requirements (C3 and C4), the resolution becomes more dominant regardless of the spatial characteristics.

Objective metrics were investigated to determine whether they yielded an OAT that correlated to that discovered using subjective testing. The results showed that there is a significant difference between the adaptation trajectories yielded using objective metrics and subjective testing techniques. This suggests that *measuring quality*

and *adapting quality based on this measurement* are different tasks.

OATS IN PRACTICE

In this section, how user-perception is affected by adapting video quality is investigated. In particular, the user-perceived quality is compared when video quality is varied by adapting the frame rate only, the resolution only, or adapting both the frame rate and the resolution using the OAT. Streaming multimedia over best-effort networks is becoming an increasingly important source of revenue. A content provider is unlikely to have the resources to provide real-time adaptive encoding for each unicast request and, as such, reserves this for "live" multicast sessions only. Typically, pre-encoded content is transmitted by unicast streams where the client chooses the connection that most closely matches their requirements. For such unicast sessions, the adaptive streaming server can employ several techniques to adapt the pre-encoded content to match the clients' resources. In such adaptive streaming systems, two techniques that are most commonly used are frame dropping and stream switching. The OAT shows how to stream the video in order to maximize the user's perceived quality in a two-dimensional adaptation space defined by frame rate and resolution (Figure 6). Adaptive frame rate can be achieved by frame dropping, while adapting spatial resolution can be achieved using track or stream switching.

All adaptation algorithms behave in an A-Increase/B-Decrease manner where A and B are the methods of change and can be either Additive, Multiplicative, Proportional, Incremental, or Decremented (Figure 7). When there is no congestion, the server increases its transmission rate either additively (AI), proportionally (PI), or multiplicatively (MI), and similarly when there is congestion, it decreases its transmission rate either additively (AD), proportionally (PD), or

Figure 6. One-dimensional versus two-dimensional adaptation

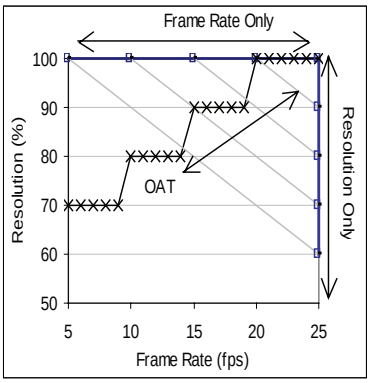
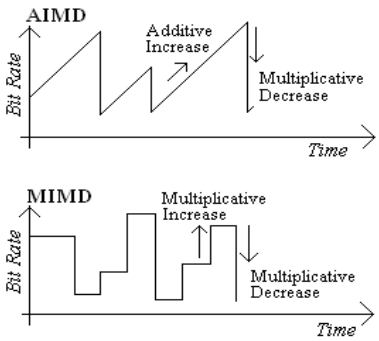


Figure 7. AIMD and MIMD



multiplicatively (MD). There are many ways to adapt video quality, for example:

- **Additive Increase/Multiplicative Decrease** (AIMD) (Chiu & Jain, 1989)
- **Additive Increase/Additive Decrease** (AIAD),
- **Additive Increase/Proportional Decrease** (AIPD) (Venkitaraman, Kim, Lee, Lu, & Bharghavan, 1999),
- **Multiplicative Increase/Multiplicative Decrease** (MIMD) (Turletti & Huitema, 1996).

In general, all rate-control algorithms exhibit some form of AI and AD behavior, although the

majority of adaptation algorithms are AIMD (Feamster, Bansal, & Balakrishnan, 2001). Thus the perception of adapting video quality is assessed in three different test cases. The first test assesses user perception when quality is adapted up in an AI manner, while the second assesses perception when quality is degraded down in an AD manner. Finally, the third assesses quality adapting in an Additive Increase/Multiplicative Decrease (AIMD) manner.

Test Methodology

The Forced Choice methodology is suitable for clips lasting not longer than 15 seconds. For video clips lasting longer than this duration, there are ***recency*** and ***forgiveness*** effects by the subject, which are a big factor when the subject must grade the overall quality of a video sequence. For example, the subject may forget and/or forgive random appearances of content-dependent artifacts when they are making their overall grade of the video sequence. To test clips of a longer duration, a different test methodology to the forced choice method needs to be applied to overcome the forgiveness and recency effects and to ensure the subject can make an accurate judgement.

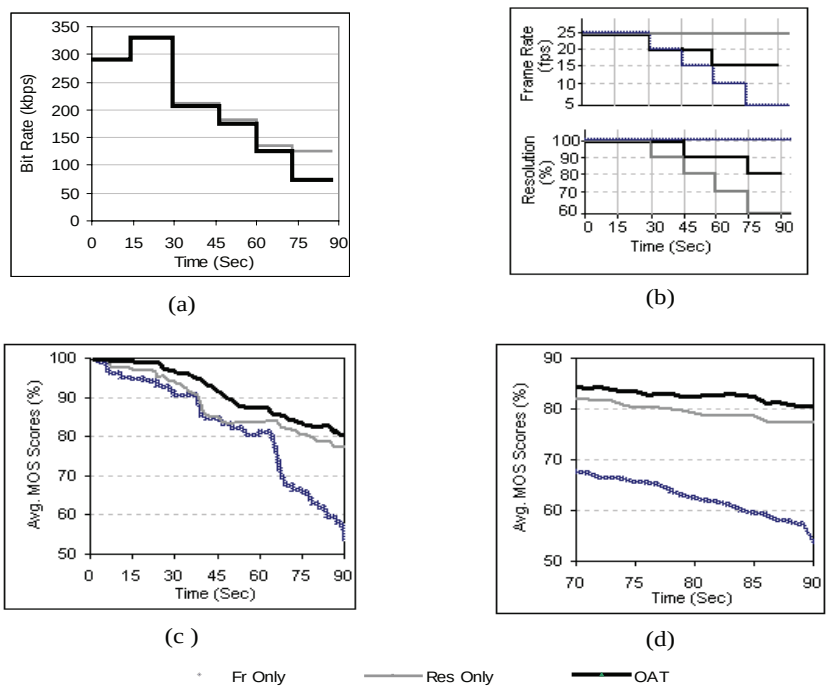
The ***Single Stimulus Continuous Quality Evaluation*** (SSCQE) methodology is intended for the presentation of sequences lasting several minutes (ITU-R BT.500-7, 1997). Continuous evaluation is performed using a slider scale on the screen to record the subjects' responses without introducing too much interference or distraction, and provides a trace of the overall quality of the sequence (Pinson & Wolf, 2003). A reference clip was played out at the beginning of the test so that the subjects were aware of the highest quality sequence. The three varying quality sequences were then presented in random order to each subject in turn. As each sequence was played out, the subject continuously rated the quality of the sequence using the slider. When the slider is moved, the quality grade of the slider is captured and related to the playout time of the media. The ***Mean Opinion Score*** (MOS) and standard deviation are calculated at each media time instant. In this case, each media time instant corresponds to one second of media. The MOS and standard deviation is calculated for each clip segment.

The test sequence chosen for this experiment contains a wide range of spatial and temporal complexity. The test sequence contains periods of high temporal complexity that are generally bursty containing many scene changes. In this test sequence, periods of high temporal complexity are generally followed by periods of relatively low temporal complexity but high spatial complexity consisting of detailed scenes such as facial close-ups and panoramic views. This test sequence contains a broad diversity of complexity and is typical of film trailers. The test sequence was divided into segments of 15 seconds duration, and each segment was encoded at various combinations of spatial resolution and frame rate. These video segments were then pieced together seamlessly to produce three varying bit rate versions of the test sequence. It was necessary to control and align each adaptation in each of the test sequences used. During these tests, it is assumed that some mechanism is implemented that informs the streaming server of the required transmission bit rate.

Results

Three scenarios were tested: First, the quality is adapted down from the best to worst; second, the quality is upgraded from worst to best; and third, the quality varies in an additive increase/multiplicative decrease fashion. The first two tests are complementary and are designed to assess symmetrical perception, that is, whether subjects perceive quality increases and quality decreases uniformly. The third test is designed to test quality perception in a typical adaptive network environment. Of particular interest are the MOS scores when the quality is decreased.

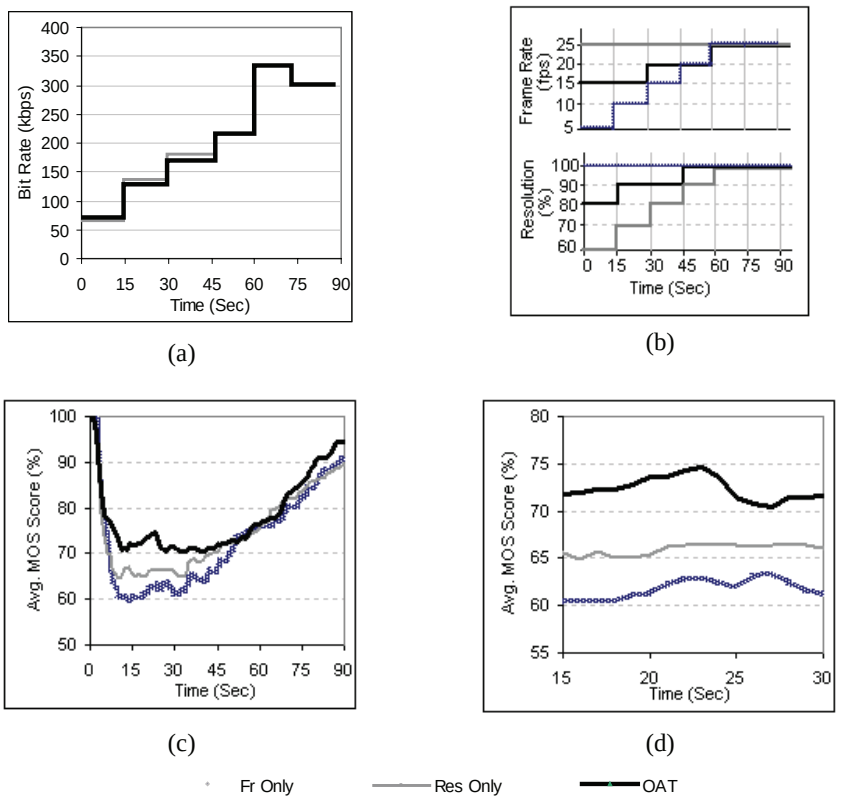
Figure 8. Time series during additive decrease (AD) in quality; (a) Segment average bit rate variations over time; (b) Video encoding parameter variations over time; (c) MOS Scores over time; (d) MOS Scores during period of lowest quality



Additive Decrease Quality Adaptation

In this test, the quality of the clip degrades from the best quality to the worst quality. Figure 8(a) shows the bit rate decreasing as the quality degrades. Figure 8(b) shows the encoding configuration of frame rate and resolution for each segment as the quality is adapting down in either the frame rate dimension only, or the resolution dimension only, or using the OAT adapting down in both the frame rate and resolution dimensions. Through time interval 0-45 seconds, the resolution and frame rate dimensions are perceived the same (Figure 8(c). In time interval 45-60 seconds, there appears to be an imperceptible difference between a decrease in resolution from 80_R to 70_R. Using the OAT, there is a smooth decrease in the MOS scores, which outperforms both one-dimensional adaptation of frame rate and resolution. During time interval 45-60 seconds, there is high action in the content which may explain the sharp decrease in the MOS scores for adapting the frame rate only. When there is high action, subjects prefer smooth continuous motion. Further, when there is high action content, reductions in spatial resolution cannot be perceived as clearly as there is too much happening in the video clip for the detail to be perceived properly. Figure 8(d) shows a close up view of MOS scores during the lowest quality level in time interval 70-90 seconds, the frame rate is perceived worst of all while the resolution performs very well. This may be due to the fact that the bit rate for the resolution is significantly greater than the two other methods. It was undesirable to achieve a lower bit rate for the resolution at 60%, as this would require a target bit rate to be set in the encoder.

Figure 9. Time series during additive increase (AI) in quality; (a) Segment average bit rate variations over time; (b) Video encoding parameter variations over time; (c) MOS Scores over time; (d) MOS Scores during period of lowest quality



Additive Increase Quality Adaptation

In this test, the quality of the clip upgrades from the worst quality to the best quality. Figure 9(b) shows the encoding configuration of frame rate and resolution as the quality is adapting up in either the frame rate dimension only or the resolution dimension only or using the OAT adapting down in both the frame rate and resolution dimensions. During this experiment, the slider is placed at the highest quality value on the rating scale when the clip begins. It can be seen that it took subjects several seconds to react to the quality level and adjust the slider to the appropriate value (Figure 9(c)). At low quality, subjects perceive adaptation using the OAT better than one-dimensional adaptation. The quality is slowly increasing, however subjects do not seem to notice the quality increasing nor do they perceive it significantly differently – indicating that subjects are more aware of quality when it is low (Figure 9(d)).

AIMD Adaptation

This section presents the results for AIMD adaptation, as might be expected from a TCP-friendly rate control mechanism. The same bit rate variation patterns were obtained in these three sequences by adapting quality in the frame rate dimension only, the spatial resolution dimension only, or both frame rate and spatial resolution dimensions, as shown in Figure 10(a). The traces in Figures 10(b) show the encoding configuration of frame

rate or resolution for each segment as the quality was adapted in either the frame rate dimension only, or the resolution dimension only, or using the OAT adapting in both the frame rate and resolution dimensions.

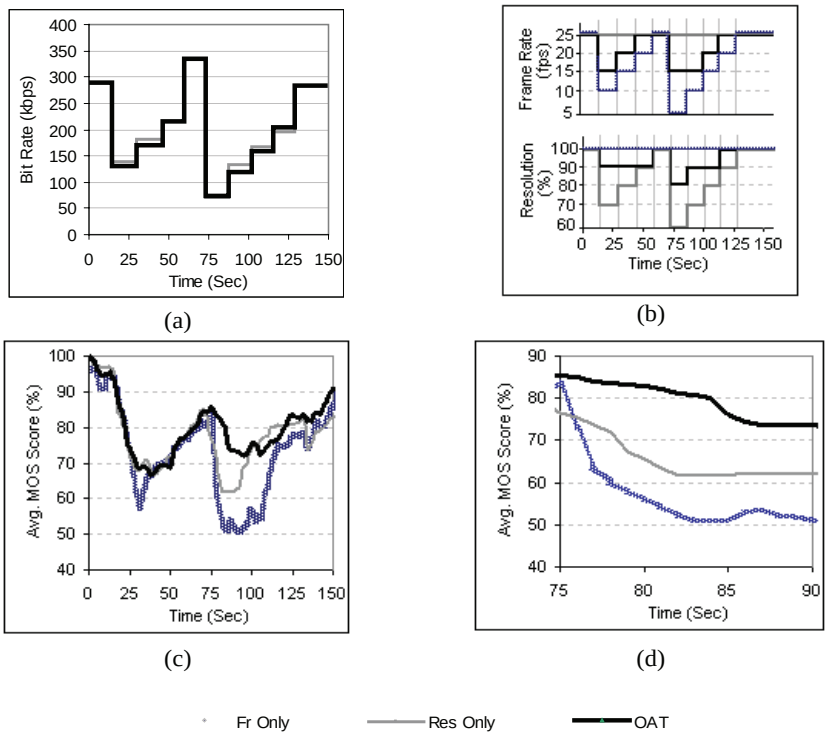
In Figure 10(a), it can be seen that although the first bit-rate reduction occurs at time 15 seconds, it is not fully perceived until time 28 seconds because there is a time delay for subjects to react to the quality adaptation. At time interval 70-90 seconds, a larger drop in bit rate occurs resulting in the lowest quality level that might reflect a mobile user entering a building. The MOS scores for adapting only the frame rate and spatial resolution are quick to reflect this drop. However, using the OAT, it takes subjects much longer to perceive this drop in quality. This is a high action part of the sequence and so the reduced frame rate is perceived more severely. The standard deviation of MOS scores using the OAT was much less than that for adapting frame rate only or spatial resolution only.

Discussion

From the experiments reported here, it appears that if a user's average bit rate changes from being quite near their maximum to near the minimum that they can tolerate, then a one-dimensional adaptation policy will cause the perceived quality to degrade quite severely. Using the two-dimensional adaptation strategy given by the OAT allows the bit rate to be dropped quite dramatically but maintain substantially better user-perceived quality.

In addition to the greater bit rate adaptation range achieved using the OAT, adaptation using the two-dimensional OAT out-performs one-dimensional adaptation using frame rate or spatial

Figure 10. Time series during additive increase multiplicative decrease (AIMD) in quality; (a) Segment average bit rate variations over time; (b) Video encoding parameter variations over time; (c) MOS Scores over time; (d) MOS Scores during period of lowest quality



resolution and reduces the variance of perception. From the various experiments conducted, subjects perceived adapting frame rate the worst, then resolution, and the OAT best of all. It was observed that there is a time delay of several seconds for subjects to react to quality adaptations. It was also observed that quality perception is asymmetrical when adapting the quality down and adapting quality up: Users are more critical of degradations in quality and less rewarding of increased quality. Similar observations were reported in Pinson and Wolf (2003).

Perception is strongly dependent on the spatio-temporal characteristics of the content. Given this understanding of user-perception, adaptation algorithms should consider the contents characteristics when making adaptation decisions. Also, frequent quality adaptation should be avoided to allow the users to become familiar with the video quality. In the experiments, the globally-averaged OAT was used, but the OAT can be dynamic if the contents' spatial and temporal characteristics are known at a given instant, thus making it more flexible to adapt according to the contents' characteristics and maximize user-perceived quality. It is expected that a dynamic OAT that adapted on the changing complexity of the content would yield even higher MOS scores.

SUMMARY

This chapter provided a brief overview of adaptive streaming systems and identified key limitations of the techniques currently in use. Quite often, adaptation algorithms omit the user-perceived quality when making adaptation decisions. Recent work in multimedia adaptation has addressed this problem by incorporating objective video quality metrics into the adaptation algorithm, thereby making the adaptation process quality-aware. However, these objective metrics have limited efficacy in assessing the user-perceived quality. As a case study, we have focused on describing recent research that attempts to address both the limitations of objective video quality metrics and adaptation techniques.

This work proposed that there is an Optimal Adaptation Trajectory (OAT), which basically states that there is an optimal way video should be adapted that maximizes the user-perceived quality. More specifically, within the set of different ways to achieve a target bit rate given by an adaptation algorithm, there exists an encoding that maximizes the user-perceived quality. Furthermore, the OAT is dependent on the spatio-temporal characteristics of the content. We have described a subjective methodology to discover the OATs through subjective testing, and applied it to finding OATs for various MPEG-4 video clips. Further it was shown that using a two-dimensional adaptation strategy given by the OAT allows the bit rate to be dropped quite dramatically but maintain substantially better user-perceived quality over one-dimensional adaptation strategies. In addition to the greater bit rate adaptation range achieved using the OAT, adaptation using the two-dimensional OAT out-performs one-dimensional adaptation using frame rate or spatial resolution and reduces the variance of perception.

Future work will assess the possibility of using and/or modifying existing objective metrics in order to mimic the OATs found by subjective methods and enable the development of a dynamic OAT. This will involve a greater analysis of the relationship between content characteristics and the corresponding OAT to determine the sensitivity of an OAT to the particular video being transmitted.

ACKNOWLEDGMENT

The support of the Research Innovation Fund and Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

REFERENCES

Chiu, D. M., & Jain, R. (1989). Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Elsevier Journal of Computer Networks and ISDN, 17*(1), 1-14.

Cranley, N., Murphy, L., & Perry, P. (2003, June). User-perceived quality aware adaptive delivery of MPEG-4 content. *Proceedings of the NOSS-DAV'03,* Monterey, California (pp. 42-49).

Feamster, N., Bansal, D., & Balakrishnan, H. (2001). On the interactions between layered quality adaptation and congestion control for streamed video. *Proceedings of Packet Video.*

Ghinea, G., Thomas, J. P., & Fish, R. S. (1999). Multimedia, network protocols, and users - Bridging the gap. *Proceedings of ACM Multimedia '99,* Orlando, Florida (pp. 473-476).

Girod, B., (1993). What's wrong with mean-squared error. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 207-220). Cambridge, MA: MIT Press.

Hekstra, A. P., Beerends, J. G., et al. (2002). PVQM - A perceptual video quality measure. *Signal Processing: Image Communication, 17*(10), 781-798.

ITU-R Recommendation BT.500-7 (1996). *Methodology for the subjective assessment of the quality of television pictures.* Geneva, Switzerland: International Telecommunication Union—Radiocommunications Sector.

ITU-T Recommendation J.143 (2000). *User requirements for objective perceptual video quality measurements in digital cable television.* Geneva, Switzerland: International Telecommunication Union—Radiocommunications Sector.

ITU-T Recommendation J.144 (2001). *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.* Geneva, Switzerland: International Telecommunication Union—Radiocommunications Sector.

ITU-T Recommendation J.148 (2003). *Requirements for an objective perceptual multimedia quality model.* Geneva, Switzerland: International Telecommunication Union—Radiocommunications Sector.

ITU-T Recommendation J.149 (2004). *Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM).* Geneva, Switzerland: International Telecommunication Union—Radiocommunications Sector.

ITU-T Recommendation P.910 (1999). *Subjective video quality assessment methods for multimedia applications.* Geneva, Switzerland: International Telecommunication Union—Radiocommunications Sector.

Masry, M., & Hemami, S. S. (2002, September). Models for the perceived quality of low bit rate video. *IEEE International Conference on Image Processing,* Rochester, NY.

Muntean, G. M., Perry, P., & Murphy, L. (2004, March). A new adaptive multimedia streaming system for all-IP multi-service networks. *IEEE Transactions on Broadcasting, 50*(1).

Pinson, M., & Wolf, S. (2003, July). Comparing subjective video quality testing methodologies. *SPIE Video Communications and Image Processing Conference,* Lugano, Switzerland.

Turletti, T., & Huitema, C. (1996). Videoconferencing on the Internet. *IEEE/ACM Transactions on Networking, 4*(3), 340-351.

van den Branden Lambrecht, C. J. (1996). Color moving pictures quality metric. *Proceedings of ICIP,* Lausanne, Switzerland (Vol. 1, pp. 885-888).

van den Branden Lambrecht, C.J., & Verscheure, O. (1996). Perceptual quality measure using a spa-

tio-temporal model of the human visual system. *Proceedings of SPIE 96,* San Jose, CA.

Venkitaraman, N., Kim, T., Lee, K. W., Lu, S., & Bharghavan, V. (1999, May). Design and evaluation of congestion control algorithms in the future Internet. *Proceedings of ACM SIGMETRICS'99,* Atlanta, Georgia.

Video Quality Experts Group (VQEG) (2005). Retrieved from http://www.its.bldrdoc.gov/vqeg/

Wang, X., & Schulzrinne, H. (1999, June). Comparison of adaptive Internet applications. *Proceedings of IEICE Transactions on Communications, E82-B*(6), 806-818.

Wang, Y., Chang, S. F., & Loui, A. (2004, June). Subjective preference of spatio-temporal rate in video adaptation using multi-dimensional scalable coding. *IEEE International Conference On Multimedia and Expo (ICME),* Taipei, Taiwan.

Watson, A., & Sasse, M. A. (1998). Measuring perceived quality of speech and video in multimedia conferencing applications. *Proceedings of ACM Multimedia '98, 12-16 September 1998,* Bristol, UK (pp. 55-60).

Watson, A. B., Hu, J., & McGowan, J. F. (2001, January). DVQ: A digital video quality metric based on human vision. *Journal of Electronic Imaging, 10*(1), 20-29.

Winkler, S. (1999). A perceptual distortion metric for digital color video. *Proceedings of the SPIE,* San Jose, CA (Vol. 3644, pp. 175-184).

Wolf, S., & Pinson, M. (1999, September 11-22). Spatial-temporal distortion metrics for in-service quality monitoring on any digital video system. *SPIE International Symposium on Voice, Video, and Data Communications,* Boston.

Wu, D., Hou, T., Zhu, W., Lee, H. J., Chiang, T., Zhang, Y.Q., & Chao, H. J. (2002). MPEG-4 video transport over the Internet: A summary. *IEEE Circuits and Systems Magazine, 2*(1), 43-46.

Wu, D., Hou, Y. T., Zhu, W., Lee, H. J., Chiang, T., Zhang, Y. Q., & Chao, H. J. (2000, September). On end-to-end architecture for transporting MPEG-4 video over the Internet. *IEEE Transactions on Circuits and Systems for Video Technology, 10*(6), 923-941.

Yu, Z., & Wu, H. R. (2000, August). Human visual system based objective digital video quality metrics. *Proceedings of the International Conference on Signal Processing of IFIP World Computer Conference 2* (pp. 1088-1095).

This work was previously published in Digital Multimedia Perception and Design, edited by G. Ghinea, and S. Y. Chen pp. 244-265, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.9
Visualization, Estimation and User Modeling for Interactive Browsing of Personal Photo Libraries

Qi Tian
University of Texas at San Antonio, USA

Baback Moghaddam
Mitsubishi Electric Research Laboratories, USA

Neal Lesh
Mitsubishi Electric Research Laboratories, USA

Chia Shen
Mitsubishi Electric Research Laboratories, USA

Thomas S. Huang
University of Illinois, USA

ABSTRACT

Recent advances in technology have made it possible to easily amass large collections of digital media. These media offer new opportunities and place great demands for new digital content user-interface and management systems which can help people construct, organize, navigate, and share digital collections in an interactive, face-to-face social setting. In this chapter, we have developed a user-centric algorithm for visualization and layout for content-based image retrieval (CBIR) in large photo libraries. Optimized layouts reflect mutual similarities as displayed on a two-dimensional (2D) screen, hence providing a perceptually intuitive visualization as compared to traditional sequential one-dimensional (1D) content-based image retrieval systems. A framework for user modeling also allows our system to learn and adapt to a user's preferences. The resulting retrieval, browsing and visualization can adapt to the user's (time-varying) notions of content, context and preferences in style and interactive navigation.

INTRODUCTION

Personal Digital Historian (PDH) Project

Recent advances in digital media technology offer opportunities for new story-sharing experiences beyond the conventional digital photo album (Balabanovic et al., 2000; Dietz & Leigh, 2001). The Personal Digital Historian (PDH) project is an ongoing effort to help people construct, organize, navigate and share digital collections in an interactive multiperson conversational setting (Shen et al., 2001; Shen et al., 2003). The research in PDH is guided by the following principles:

1. The display device should enable natural face-to-face conversation: not forcing everyone to face in the same direction (desktop) or at their own separate displays (hand-held devices).
2. The physical sharing device must be convenient and customary to use: helping to make the computer disappear.
3. Easy and fun to use across generations of users: minimizing time spent typing or formulating queries.
4. Enabling interactive and exploratory storytelling: blending authoring and presentation.

Current software and hardware do not meet our requirements. Most existing software in this area provides users with either powerful query methods or authoring tools. In the former case, the users can repeatedly query their collections of digital content to retrieve information to show someone (Kang & Shneiderman, 2000). In the latter case, a user experienced in the use of the authoring tool can carefully craft a story out of his or her digital content to show or send to someone at a later time. Furthermore, current hardware is also lacking. Desktop computers are not suitably designed for group, face-to-face conversation in a social setting, and handheld story-telling devices have limited screen sizes and can be used only by a small number of people at once. The objective of the PDH project is to take a step beyond.

The goal of PDH is to provide a new digital content user-interface and management system enabling face-to-face casual exploration and visualization of digital contents. Unlike conventional desktop user interface, PDH is intended for multiuser collaborative applications on single display groupware. PDH enables casual and exploratory retrieval, and interaction with and visualization of digital contents.

We design our system to work on a touch-sensitive, circular tabletop display (Vernier et al., 2002), as shown in Figure 1. The physical PDH table that we use is a standard tabletop with a top projection (either ceiling mounted or tripod mounted) that displays on a standard whiteboard as shown in the right image of Figure 1. We use two Mimio (www.mimio.com/meet/mimiomouse) styluses as the input devices for the first set of user experiments. The layout of the entire tabletop display consists of (1) a large story-space area encompassing most of the tabletop until the perimeter, and (2) one or more narrow arched control panels (Shen et al., 2001). Currently, the present PDH table is implemented using our DiamondSpin (www.merl.com/projects/diamondspin) circular table Java toolkit. DiamondSpin is intended for multiuser collaborative applications (Shen et al., 2001; Shen et al., 2003; Vernier et al., 2002).

The conceptual model of PDH is to focus on developing content organization and retrieval metaphors that can be easily comprehended by users without distracting from the conversation. We adopt a model of organizing the materials using the four questions essential to storytelling: who, when, where, and what (the four Ws). We do not currently support why, which is also useful for storytelling. Control panels located on the perimeter of the table contain buttons labeled "people," "calendar," "location," and "events," corresponding to these four questions. When a

user presses the "location" button, for example, the display on the table changes to show a map of the world. Every picture in the database that is annotated with a location will appear as a tiny thumbnail at its location. The user can pan and zoom in on the map to a region of interest, which increases the size of the thumbnails. Similarly, by pressing one of the other three buttons, the user can cause the pictures to be organized by the time they were taken along a linear timeline, the people they contain, or the event keywords with which the pictures were annotated. We assume the pictures are partially annotated. Figure 2 shows an example of navigation of a personal photo album by the four-Ws model. Adopting this model allows users to think of their documents in terms of how they would like to record them as part of their history collection, not necessarily in a specific hierarchical structure. The user can make selections among the four Ws and PDH will automatically combine them to form rich Boolean queries implicitly for the user (Shen et al., 2001; Shen, Lesh, Vernier, Forlines, & Frost, 2002; Shen et al., 2003; Vernier et al., 2002).

The PDH project combines and extends research in largely two areas: (i) human-computer interaction (HCI) and interface (the design of the shared-display devices, user interface for storytelling and online authoring, and storylistening) (Shen et al., 2001, 2002, 2003; Vernier et al., 2002); (ii) content-based information visualization, presentation and retrieval (user-guided image layout, data mining and summarization) (Moghaddam et al., 2001, 2002, 2004; Tian et al., 2001, 2002). Our work has been done along these two lines. The work by Shen et al. (2001, 2002, 2003) and Vernier et al. (2002) focused on the HCI and interface design issue of the first research area. The work in this chapter is under the context of PDH but focuses on the visualization, smart layout, user modeling and retrieval part. In this chapter, we propose a novel visualization and layout algorithm that can enhance informal storytelling using personal digital data such as photos, audio and video in a face-to-face social setting. A framework for user modeling also allows our system to learn and adapt to a user's preferences. The resulting retrieval, browsing and visualization can adapt to the user's (time-varying) notions of content, context and preferences in style and interactive navigation.

Related Work

In content-based image retrieval (CBIR), most current techniques are restricted to matching image appearance using primitive features such as color, texture, and shape. Most users wish to retrieve images by semantic content (the objects/events depicted) rather than by appearance. The resultant *semantic gap* between user expectations and the current technology is the prime cause of the poor takeup of CBIR technology. Due to the semantic gap (Smeulders et al., 2000), visualization becomes very important for user to navigate the complex query space. New visualization tools are required to allow for user-dependent and goal-dependent choices about what to display and how to provide feedback. The query result has an inherent display dimension that is often ignored. Most methods display images in a 1D list in order of decreasing similarity to the query images. Enhancing the visualization of the query results is, however, a valuable tool in helping the user navigate query space. Recently, Horoike and Musha (2000), Nakazato and Huang (2001), Santini and Jain (2000), Santini et al. (2001), and Rubner (1999) have also explored toward content-based visualization. A common observation in these works is that the images are displayed in 2D or 3D space from the projection of the high-dimensional feature spaces. Images are placed in such a way that distances between images in 2D or 3D reflect their distances in the high-dimensional feature space. In the works of Horoike and Musha (2000) and Nakazato and Huang (2001), the users can view large sets of images in 2D or 3D space and user navigation is allowed. In the

Figure 1. PDH table (a) an artistic rendering of the PDH table (designed by Ryan Bardsley, Tixel HCI www.tixel.net) and (b) the physical PDH table

(a)

(b)

works of Nakazato and Huang (2001), Santini et al. (2000, 2001), the system allows user interaction on image location and forming new groups. In the work of Santini et al. (2000, 2001), users can manipulate the projected distances between images and learn from such a display.

Our work (e.g., Tian et al., 2001, 2002; Moghaddam et al., 2001, 2002, 2004) under the context of PDH shares many common features with the related work (Horoike & Musha, 2000; Nakazato & Huang, 2001; Santini et al., 2000, 2001; Rubner, 1999). However, a learning mechanism from the display is not implemented in Horoike and Musha (2000), and 3D MARS (Nakazato & Huang, 2001) is an extension to our work (Tian et al., 2001; Moghaddam et al. 2001) from 2D to 3D space. Our system differs from the work of Rubner (1999) in that we adopted different mapping methods. Our work shares some features with the work by Santini and Jain (2000) and Santini et al. (2001) except that our PDH system is currently being incorporated into a much broader system for computer human-guided navigating, browsing, archiving, and interactive storytelling with large

Figure 2. An example of navigation by the four-Ws model (Who, When, Where, What)

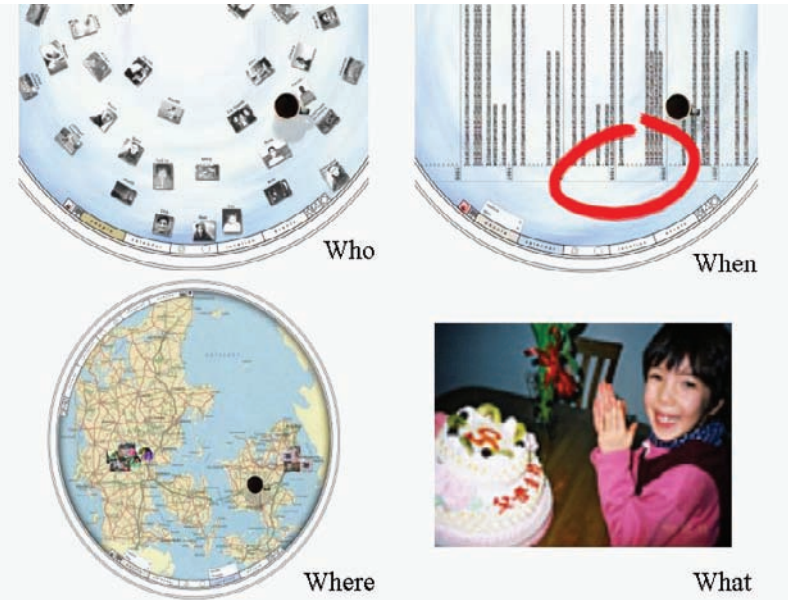


photo libraries. The part of this system described in the remainder of this chapter is, however, specifically geared towards adaptive user modeling and relevance estimation and based primarily on visual features as opposed to semantic annotation as in Santini and Jain (2000) and Santini et al. (2001).

The rest of the chapter is organized as follows. In Content-Based Visualization, we present designs for uncluttered visualization and layout of images (or iconic data in general) in a 2D display space for content-based image retrieval (Tian et al., 2001; Moghaddam et al., 2001). In Context and User Modeling, we further provide a mathematical framework for user modeling, which adapts and mimics the user's (possibly changing) preferences and style for interaction, visualization and navigation (Moghaddam et al., 2002, 2004; Tian et al., 2002). Monte Carlo simulations in the Statistical Analysis section plus the next section on User Preference Study have demonstrated the ability of our framework to model or "mimic" users, by automatically generating layouts according to user's preference. Finally, Discussion and Future Work are given in the final section.

CONTENT-BASED VISUALIZATION

With the advances in technology to capture, generate, transmit and store large amounts of digital imagery and video, research in content-based image retrieval (CBIR) has gained increasing attention. In CBIR, images are indexed by their visual contents such as color, texture, and so forth. Many research efforts have addressed how to extract these low-level features (Stricker & Orengo, 1995; Smith & Chang, 1994; Zhou et al., 1999), evaluate distance metrics (Santini & Jain, 1999; Popescu & Gader, 1998) for similarity measures and look for efficient searching schemes (Squire et al. 1999; Swets & Weng, 1999).

In this section, we present a user-centric algorithm for visualization and layout for content-based image retrieval. Image features (visual and/or semantic) are used to display retrievals as

thumbnails in a 2D spatial layout or "configuration" which conveys pair-wise mutual similarities. A graphical optimization technique is used to provide maximally uncluttered and informative layouts. We should note that one physical instantiation of the PDH table is that of a roundtable, for which we have in fact experimented with polar coordinate conformal mappings for converting traditional rectangular display screens. However, in the remainder of this chapter, for purposes of ease of illustration and clarity, all layouts and visualizations are shown on rectangular displays only.

Traditional Interfaces

The purpose of automatic content-based visualization is augmenting the user's understanding of large information spaces that cannot be perceived by traditional sequential display (e.g., by rank order of visual similarities). The standard and commercially prevalent image management and browsing tools currently available primarily use tiled sequential displays — *that is,* essentially a simple 1D similarity based visualization.

However, the user quite often can benefit by having a global view of a working subset of retrieved images in a way that reflects the relations between *all pairs* of images — that is, N^2 measurements as opposed to only N. Moreover, even a narrow view of one's immediate surroundings defines "context" and can offer an indication on how to explore the dataset. The wider this "visible" horizon, the more efficient the new query will be formed. Rubner (1999) proposed a 2D display technique based on multidimensional scaling (MDS) (Torgeson, 1998). A global 2D view of the images is achieved that reflects the mutual similarities among the retrieved images. MDS is a nonlinear transformation that minimizes the stress between high-dimensional feature space and low-dimensional display space. However, MDS is rotation invariant, nonrepeatable (nonunique), and often slow to implement. Most critically, MDS (as well as some of the other leading nonlinear dimensionality reduction methods) provide high-to-low-dimensional projection operators that are not analytic or functional in form, but are rather defined on a point-by-point basis for each given dataset. This makes it very difficult to project a new dataset in a functionally consistent way (without having to build a post-hoc projection or interpolation function for the forward mapping each time). We feel that these drawbacks make MDS (and other nonlinear methods) an unattractive option for real-time browsing and visualization of high-dimensional data such as images.

Improved Layout and Visualization

We propose an alternative 2D display scheme based on Principal Component Analysis (PCA) (Jolliffe, 1996). Moreover, a novel window display optimization technique is proposed which provides a more perceptually intuitive, visually uncluttered and informative visualization of the retrieved images.

Traditional image retrieval systems display the returned images as a list, sorted by decreasing similarity to the query. The traditional display has one major drawback. The images are ranked by similarity to the query, and relevant images (as for example used in a relevance feedback scenario) can appear at separate and distant locations in the list. We propose an alternative technique to MDS (Torgeson, 1998) that displays mutual similarities on a 2D screen based on visual features extracted from images. The retrieved images are displayed not only in ranked order of similarity from the query but also according to their mutual similarities, so that similar images are grouped together rather than being scattered along the entire returned 1D list.

Visual Features

We will first describe the low-level visual feature extraction used in our system. There are three

visual features used in our system: color moments (Stricker & Orengo, 1995), wavelet-based texture (Smith & Chang, 1994), and water-filling edge-based structure feature (Zhou et al., 1999).

The color space we use is HSV because of its decorrelated coordinates and its perceptual uniformity (Stricker & Orengo, 1995). We extract the first three moments (mean, standard deviation and skewness) from the three-color channels and therefore have a color feature vector of length $3\times3 = 9$.

For wavelet-based texture, the original image is fed into a wavelet filter bank and is decomposed into 10 decorrelated subbands. Each subband captures the characteristics of a certain scale and orientation of the original image. For each subband, we extract the standard deviation of the wavelet coefficients and therefore have a texture feature vector of length 10.

For water-filling edge-based structure feature vector, we first pass the original images through an edge detector to generate their corresponding edge map. We extract eighteen (18) elements from the edge maps, including *max fill time, max fork count, and so forth*. For a complete description of this edge feature vector, interested readers are referred to Zhou et al. (1999).

Dimension Reduction and PCA Splats

To create such a 2D layout, Principal Component Analysis (PCA) (Jolliffe, 1996) is first performed on the retrieved images to project the images from the high-dimensional feature space to the 2D screen. Image thumbnails are placed on the screen so that the screen distances reflect as closely as possible the similarities between the images. If the computed similarities from the high-dimensional feature space agree with our perception, and if the resulting feature dimension reduction preserves these similarities reasonably well, then the resulting spatial display should be informative and useful.

In our experiments, the 37 visual features (nine color moments, 10 wavelet moments and 18 water-filling features) are preextracted from the image database and stored off-line. Any 37-dimensional feature vector for an image, when taken in context with other images, can be projected onto the 2D screen based on the first two principal components normalized by the respective eigenvalues. Such a layout is denoted as a PCA Splat. We implemented both linear and nonlinear projection methods using PCA and Kruskal's algorithm (Torgeson, 1998). The projection using the nonlinear method such as the Kruskal's algorithm is an iterative procedure, slow to converge and converged to the local minima. Therefore the convergence largely depends on the initial starting point and cannot be repeated. On the contrary, PCA has several advantages over nonlinear methods like MDS. It is a fast, efficient and unique linear transformation that achieves the maximum distance preservation from the original high-dimensional feature space to 2D space among all possible linear transformations (Jolliffe, 1996). The fact that it fails to model nonlinear mappings (which MDS succeeds at) is in our opinion a minor compromise given the advantages of real-time, repeatable and mathematically tractable linear projections.

We should add that nonlinear dimensionality reduction (NLDR) by itself is a very large area of research and mostly beyond the scope of this chapter. We only comment on MDS because of its previous use by Rubner (1999) for CBIR. Use of other iterative NLDR techniques as "principal curves" or "bottleneck" auto-associative feedforward networks is usually prohibited by the need to perform real-time and repeatable projections. More recent advances such as IsoMap (Tennenbaum et al., 2000) and Local Linear Embedding (LLE) (Roweis & Saul, 2000) are also not amendable to real-time or closed-form computation. The most recent techniques such as Laplacian Eigenmaps (Belkin & Niyogi, 2003) and charting (Brand, 2003) have only just begun to be used and may promise advances useful in this

Figure 3. Top 20 retrieved images (ranked top to bottom and left to right; query is shown first in the list)

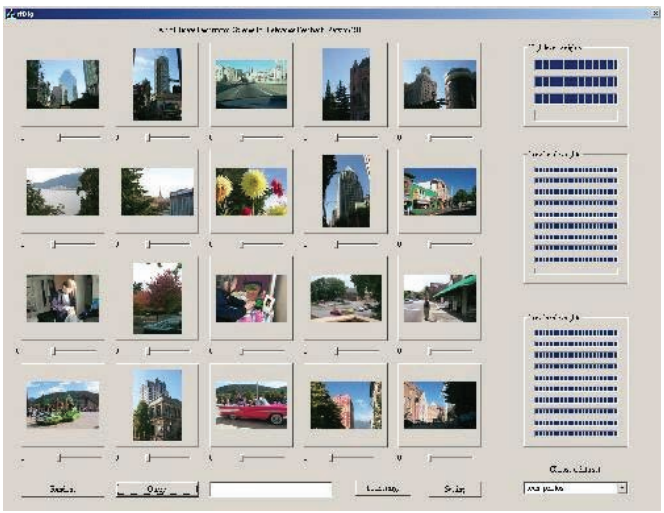
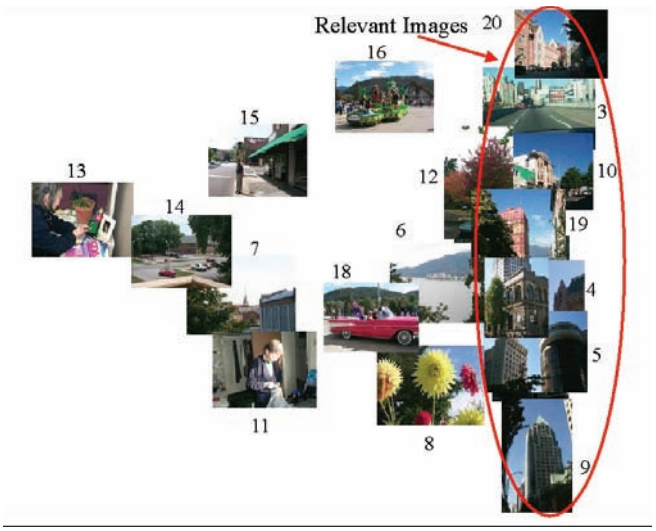


Figure 4. PCA Splat of top 20 retrieved images in Figure 3



application domain, although we should hasten to add that the formulation of subspace weights and their estimation (see section on Context and User Modeling) is not as straightforward as with the case of linear dimension reduction (LDR) methods like PCA.

Let us consider a scenario of a typical image-retrieval engine at work in which an actual user is providing relevance feedback for the purposes of query refinement. Figure 3 shows an example of the retrieved images by the system (which resembles most traditional browsers in its 1D tile-based layout). The database is a collection of 534 images. The first image (building) is the query. The other nine relevant images are ranked in second, third, fourth, fifth, ninth, 10th, 17th, 19th and 20th places, respectively.

Figure 4 shows an example of a PCA Splat for the top 20 retrieved images shown in Figure 3. In addition to visualization by layout, in this particular example, the sizes (alternatively contrast) of the images are determined by their visual similarity to the query. The higher the rank, the larger is the size (or higher the contrast). There is also a number next to each image in Figure 4 indicating its corresponding rank in Figure 4. The view of query image, that is, the top left one in Figure 3, is blocked by the images ranked 19th, fourth, and 17th in Figure 4. A better view is achieved in Figure 7 after display optimization.

Clearly the relevant images are now better clustered in this new layout as opposed to being dispersed along the tiled 1D display in Figure 3. Additionally, PCA Splats convey N^2 mutual distance measures relating all pair-wise similarities between images, whereas the ranked 1D display in Figure 3 provides only N.

Display Optimization

However, one drawback of PCA Splat is that some images can be partially or totally overlapped, which makes it difficult to view all the images at the same time. The overlap will be even worse when the number of retrieved images becomes larger, for example, larger than 50. To solve the overlapping problem between the retrieved images, a novel optimized technique is proposed in this section.

Given a set the retrieved images and their corresponding sizes and positions, our optimizer tries to find a solution that places the images at the appropriate positions while deviating as little as possible from their initial PCA Splat positions. Assume the number of images is N. The image positions are represented by their center coordinates (x_i, y_i), $i = 1, ..., N$, and the initial image positions are denoted as (x^o_i, y^o_i), $i = 1, ..., N$. The minimum and maximum coordinates of the 2D screen are $[x_{min}, x_{max}, y_{min}, y_{max}]$. The image size is represented by its radius r_i for simplicity, $i = 1, ..., N$ and the maximum and minimum image size is r_{max} and r_{min} in radius, respectively. The initial image size is r^o_i, $i = 1, ..., N$.

To minimize the overlap, the images can be automatically moved away from each other to decrease the overlap between images, but this will increase the deviation of the images from their initial positions. Large deviation is certainly undesirable because the initial positions provide important information about mutual similarities between images. So there is a trade-off problem between minimizing overlap and minimizing deviation. Without increasing the overall deviation, an alternative way to minimize the overlap is to simply shrink the image size as needed, down to a minimum size limit. The image size will not be increased in the optimization process because this will always increase the overlap. For this reason, the initial image size r^o_i is assumed to be r_{max}.

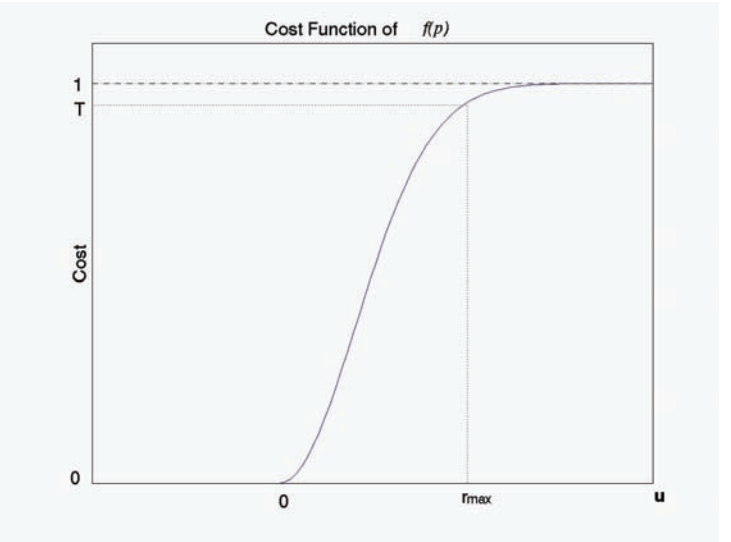
The total cost function is designed as a linear combination of the individual cost functions taking into account two factors. The first factor is to keep the overall overlap between the images on the screen as small as possible. The second factor is to keep the overall deviation from the initial position as small as possible.

$$J = F(p) + \lambda \cdot S \cdot G(p) \qquad (1)$$

where $F(p)$ is the cost function of the overall overlap and $G(p)$ is the cost function of the overall deviation from the initial image positions, S is a scaling factor which brings the range of $G(p)$ to the same range of $F(p)$, and S is chosen to be $(N–1)/2$. λ is a weight and $\lambda \geq 0$. When λ is zero, the deviation of images is not considered in overlapping minimization. When λ is less than one, minimizing overall overlap is more important than minimizing overall deviation, and vice versa for λ is greater than one.

The cost function of overall overlap is designed as

Figure 5. Cost function of overlap function f(p)



$$F(p) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(p) \tag{2}$$

$$f(p) = \begin{cases} 1 - e^{-\frac{u^2}{\sigma_f}} & u > 0 \\ 0 & u \le 0 \end{cases} \tag{3}$$

where $u = r_i + r_j - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, is a measure of overlapping. When $u \le 0$, there is no overlap between the i^{th} image and the j^{th} image, thus the cost is 0. When $u > 0$, there is partial overlap between the i^{th} image and the j^{th} image. When $u = 2 \cdot r_{max}$, the i^{th} image and the j^{th} image are totally overlapped. σ_f is a curvature-controlling factor.

Figure 5 shows the plot of $f(p)$. With the increasing value of $u (u > 0)$, the cost of overlap is also increasing.

From Figure 5, in Equation (3) is calculated by setting $T=0.95$ when $u = r_{max}$.

$$\sigma_f = \frac{-u^2}{\ln(1-T)}\Big|_{u=r_{\max}} \tag{4}$$

The cost function of overall deviation is designed as

$$G(p) = \sum_{i=1}^{N} g(p) \tag{5}$$

$$g(p) = 1 - e^{-\frac{v^2}{\sigma_g}} \tag{6}$$

where $v = \sqrt{(x_i - x_i^o)^2 + (y_i - y_i^o)^2}$, v is the measure of deviation of the i^{th} image from its initial position. σ_g is a curvature-controlling factor. (x_i, y_i) and (x^o_i, y^o_i) are the optimized and initial center coordinates of the i^{th} image, respectively, i = 1, ..., N.

Figure 6 shows the plot of $g(p)$. With the increasing value of v, the cost of deviation is also increasing.

From Figure 6, σ_g in Equation (6) is calculated by setting $T=0.95$ when $v = maxsep$. In our work, $maxsep$ is set to be $2 \cdot r_{max}$.

$$\sigma_g = \frac{-v^2}{\ln(1-T)}\Big|_{v=maxsep} \tag{7}$$

The optimization process is to minimize the total cost J by finding a (locally) optimal set of size and image positions. The nonlinear optimization method was implemented by an iterative gradient descent method (with line search). Once converged, the images will be redisplayed based on the new optimized sizes and positions.

Figure 6. Cost function of function g(p)

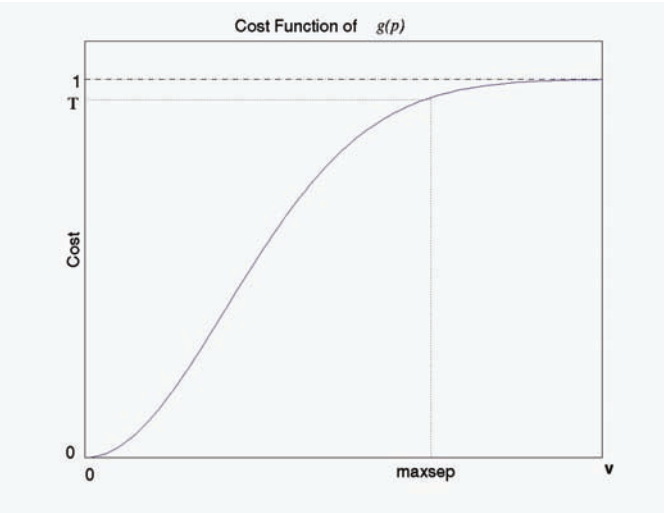


Figure 7. Optimized PCA Splat of Figure 3

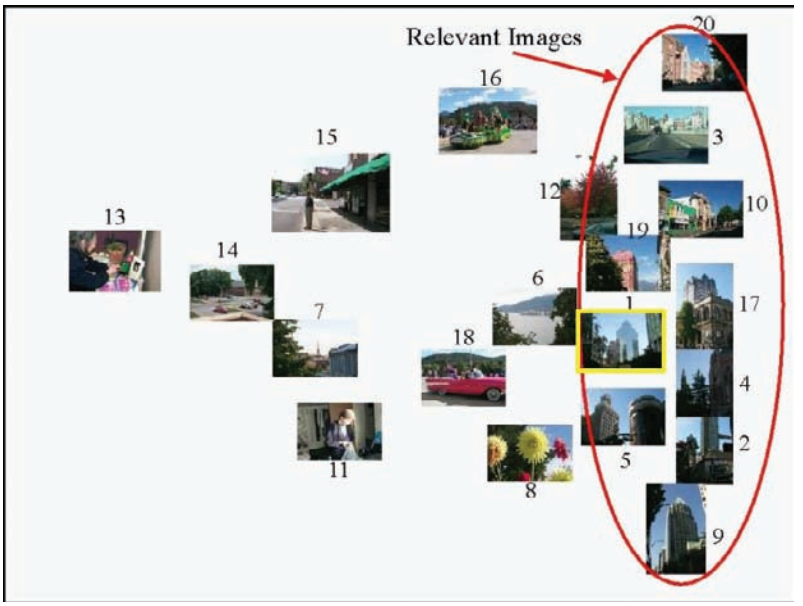


Figure 7 shows the optimized PCA Splats for Figure 3. The image with a yellow frame is the query image in Figure 3. Clearly, the overlap is minimized while the relevant images are still close to each other to allow a global view. With such a display, the user can see the relations between the images, better understand how the query performed, and subsequently formulate future queries more naturally. Additionally, attributes such as contrast and brightness can be used to convey rank. We note that this additional visual aid is essentially a third dimension of information display. For example, images with higher rank could be displayed with larger size or increased

brightness to make them stand out from the rest of the layout. An interesting example is to display "time" or "timeliness" by associating the size or brightness with how long ago the picture was taken, thus images from the "past" would appear smaller or dimmer than those taken recently. A full discussion of the resulting enhanced layouts is deferred to future work.

Also we should point out that despite our ability to "clean-up" layouts for maximal visibility with the optimizer we have designed, all subsequent figures in this chapter show Splats *without* any overlap minimization, because, for illustrating (as well as comparing) the accuracy of the estimation results in subsequent sections, the absolute position was necessary and important.

CONTEXT AND USER MODELING

Image content and "meaning" is ultimately based on semantics. The user's notion of content is a high-level concept, which is quite often removed by many layers of abstraction from simple low-level visual features. Even near-exhaustive semantic (keyword) annotations can never fully capture context-dependent notions of content. The same image can "mean" a number of different things depending on the particular circumstance. The visualization and browsing operation should be aware of which features (visual and/or semantic) are relevant to the user's current focus (or working set) and which should be ignored. In the space of all possible features for an image, this problem can be formulated as a subspace identification or feature weighting technique that is described fully in this section.

Estimation of Feature Weights

By user modeling or "context awareness" we mean that our system must be constantly aware of and adapting to the changing concepts and preferences of the user. A typical example of this human-computer synergy is having the system learn from a user-generated layout in order to visualize new examples based on identified relevant/irrelevant features. In other words, design smart browsers that "mimic" the user, and over time, adapt to their style or preference for browsing and query display. Given information from the layout, for example, positions and mutual distances between images, a novel feature weight estimation scheme, noted as α-estimation is proposed, where α is a weighting vector for features, for example, color, texture and structure (and semantic keywords).

We now describe the subspace estimation of α for visual features only, for example, color, texture, and structure, although it should be understood that the features could include visual, audio and semantic features or any hybrid combination thereof.

In theory, the estimation of weights can be done for all the visual features if given enough images in the layout. The mathematical formulation of this estimation problem follows.

The weighing vector is $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_L\}$, where L is the total length of color, texture, and structure feature vector, for example, $L = 37$ in this chapter. The number of images in the preferred clustering is N, and X is an $L \times N$ matrix where the i^{th} column is the feature vector of the i^{th} image, i, $j = 1, ..., N$. The distance, for example Euclidean-based between the i^{th} image and the j^{th} image, for $i, j = 1, ..., N$, in the preferred clustering (distance in 2D space) is d_{ij}. These weights $\alpha_1, \alpha_2, ..., \alpha_L$ are constrained such that they always sum to 1.

We then define an energy term to minimize with an L_p norm (with $p = 2$). This cost function is defined in Equation (8). It is a nonnegative quantity that indicates how well mutual distances are preserved in going from the original high-dimensional feature space to 2D space. Note that this cost function is similar to MDS stress, but unlike MDS, the minimization is seeking the optimal feature weights α. Moreover, the low-dimensional projections in this case are already known. The optimal weighting parameter recovered is then

used to weight original feature-vectors before applying a PCA Splat which will result in the desired layout.

$$J = \sum_{i=1}^{N}\sum_{j=1}^{N}(d_{ij}{}^{p} - \sum_{k=1}^{L}\alpha_k{}^{p} \mid \mathbf{X}_i{}^{(k)} - \mathbf{X}_j{}^{(k)} \mid^{p})^2 \quad (8)$$

The global minimum of this cost function corresponding to the optimal weight parameter α, is easily obtained using a constrained (non-negative) least-squares. To minimize *J*, take the partial derivative of *J* relative to for $l = 1, ..., L$ and set them to zero, respectively.

$$\frac{\partial J}{\partial \alpha_l{}^{p}} = 0 \qquad l = 1, \cdots, L \quad (9)$$

We thus have

$$\sum_{k=1}^{L}\alpha_k{}^{p}\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{X}_i{}^{(l)} - \mathbf{X}_j{}^{(l)}|^{p}|\mathbf{X}_i{}^{(k)} - \mathbf{X}_j{}^{(k)}|^{p} = \sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}{}^{p} \mid \mathbf{X}_i{}^{(l)} - \mathbf{X}_j{}^{(l)}|^{p}$$

$$l = 1, \cdots, L \quad (10)$$

Define

$$R(l, k) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mid \mathbf{X}_i{}^{(l)} - \mathbf{X}_j{}^{(l)} \mid^{p} \mid \mathbf{X}_i{}^{(k)} - \mathbf{X}_j{}^{(k)} \mid^{p} \quad (11)$$

$$r(l) = \sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}{}^{p} \mid \mathbf{X}_i{}^{(l)} - \mathbf{X}_j{}^{(l)} \mid^{p} \quad (12)$$

and subsequently simplify Equation (10) to:

$$\sum_{k=1}^{L}\alpha_k{}^{p} R(l, k) = r(l) \quad l = 1, \cdots, L \quad (13)$$

Using the following matrix/vector definitions

$$R = \begin{bmatrix} R(1,1) & R(1,2) & \cdots & R(1, L) \\ R(2,1) & R(2,2) & \cdots & R(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ R(L,1) & R(L,2) & \cdots & R(L,L) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \alpha_1{}^{p} \\ \alpha_2{}^{p} \\ \vdots \\ \alpha_L{}^{p} \end{bmatrix}$$

and

$$r = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(L) \end{bmatrix}$$

Equation (13) is simplified to

$$R \cdot \beta = r \quad (14)$$

Subsequently β is obtained as a constrained ($\beta > 0$) linear least-squares solution of the above system. The weighting vector α_k is then simply determined by the *p-th* root of β where we typically use *p=2.*

We note that there is an alternative approach to estimating the subspace weighting vector α in the sense of *minimum deviation*, which we have called "deviation-based" α-estimation. The cost function in this case is defined as follows:

$$J = \sum_{i=1}^{N} \mid p^{(i)}(x, y) - \hat{p}^{(i)}(x, y) \mid^{2} \quad (15)$$

where $p^{(i)}(x,y)$ and are the original and projected 2D locations of the i^{th} image, respectively. This formulation is a more direct approach to estimation since it deals with the final position of the images in the layout. Unfortunately, however, this approach requires the simultaneous estimation of both the weight vectors as well as the projection basis and consequently requires less-accurate iterative *re-estimation* techniques (as opposed to more robust closed-form solutions possible with Equation (8)). A full derivation of the solution for our deviation-based α estimation is shown in Appendix A.

Compare two different estimation methods: stress-based and deviation-based. The former is most useful, robust, and "identifiable" in the control theory sense of the word. The latter uses a somewhat unstable re-estimation framework and does not always give satisfactory results. However, we still provide a detailed description for the sake of completeness. The shortcomings of this latter method are immediately apparent from the solution requirements. This discussion can be found in Appendix A.

For the reasons mentioned above, in all the experiments reported in this chapter, we use only the stress-based method of Equation (8) for α estimation.

We note that in principle it is possible to use a single weight for each dimension of the feature vector. However, this would lead to a poorly determined estimation problem since it is unlikely (and/or undesirable) to have that many sample images from which to estimate all individual weights. Even with plenty of examples (an over-determined system), chances are that the estimated weights would generalize poorly to a new set of images — this is the same principle used in a modeling or regression problem where the order of the model or number of free parameters should be less than the number of available observations.

Therefore, in order to avoid the problem of *over-fitting* and the subsequent poor generalization on new data, it is ideal to use fewer weights. In this respect, the less weights (or more subspace "groupings") there are, the better the generalization performance. Since the origin of all visual features, that is, 37 features, is basically from three different (independent) visual attributes: color, texture and structure, it seems prudent to use three weights corresponding to these three subspaces. Furthermore, this number is sufficiently small to almost guarantee that we will always have enough images in one layout from which to estimate these three weights. Therefore, in the remaining portion of the chapter, we only estimated a weighting vector , where α_c is the weight for color feature of length L_c, α_t is the weight for texture feature of length L_t, α_s and is the weight for structure feature of length , respectively. These weights α_c, α_t, α_s are constrained such that they always sum to 1, and $L = L_c + L_t + L_s$.

Figure 8 shows a simple user layout where three car images are clustered together despite their different colors. The same is performed with three flower images (despite their texture/structure). These two clusters maintain a sizeable separation, thus suggesting two separate concept classes implicit by the user's placement. Specifically, in this layout the user is clearly concerned with the distinction between *car* and *flower* regardless of color or other possible visual attributes.

Applying the α-estimation algorithm to Figure 8, the feature weights learned from this layout are $\alpha_c = 0.3729$, $\alpha_t = 0.5269$ and $\alpha_s = 0.1002$. This

Figure 8. An example of a user-guided layout

Figure 9. PCA Splat on a larger set of images using (a) estimated weights (b) arbitrary weights

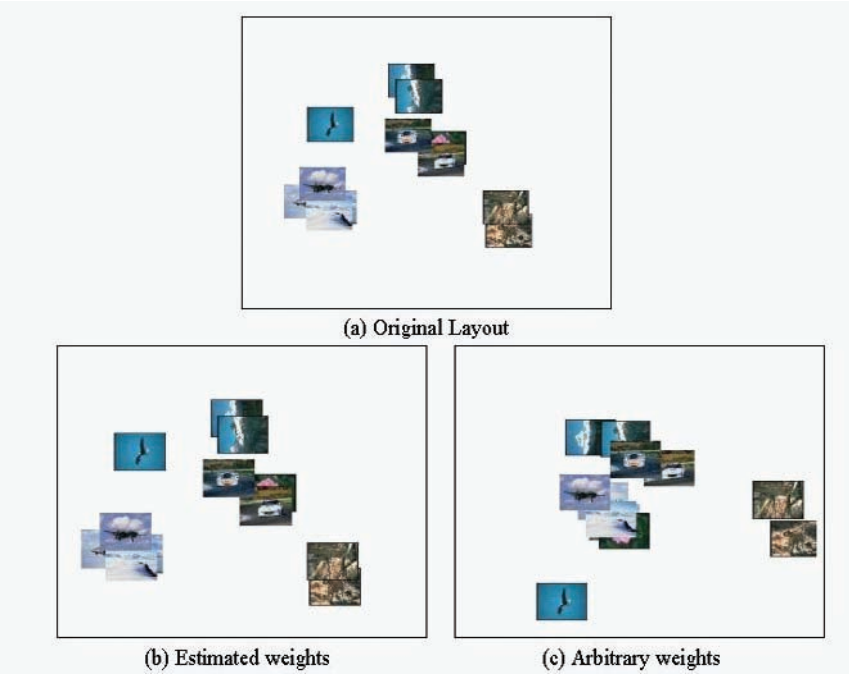
(a) *(b)*

shows that the most important feature in this case is texture and not color, which is in accord with the concepts of car versus flower as graphically indicated by the user in Figure 8.

Now that we have the learned feature weights (or modeled the user) what can we do with them? Figure 9 shows an example of a typical application: automatic layout of a larger (more complete data set) set of images in the style indicated by the user. Figure 9(a) shows the PCA Splat using the learned feature weight for 18 cars and 19 flowers. It is obvious that the PCA Splat using the estimated weights captures the essence of the configuration layout in Figure 8. Figure 9(b) shows a PCA Splat of the same images but with a randomly generated α, denoting an arbitrary but coherent 2D layout, which in this case, favors color ($\alpha_c = 0.7629$). This comparison reveals that proper feature weighting is an important factor in generating the user-desired and sensible layouts. We should point out that a random α does not generate a random layout, but rather one that is still coherent, displaying consistent groupings or clustering. Here we have used such "random" layouts as substitutes for alternative (arbitrary) layouts that are nevertheless "valid" (differing only in the relative contribution of the three features to the final design of the layout). Given the difficulty of obtaining hundreds (let alone thousands) of real user layouts that are needed for more complete statistical tests (such as those in the next section), random α layouts are the only conceivable way of "simulating" a layout by a real user in accordance with "familiar" visual criteria such as color, texture or structure.

Figure 10(a) shows an example of another layout. Figure 10(b) shows the corresponding computer-generated layout of the same images with their high-dimensional feature vectors weighted by the estimated α, which is recovered solely from the 2D configuration of Figure 10(a). In this instance the reconstruction of the layout is near perfect, thus demonstrating that our high-dimensional subspace feature weights can in fact

Figure 10. (a) An example layout. Computer-generated layout based on (b) reconstruction using learned feature weights, and (c) the control (arbitrary weights)



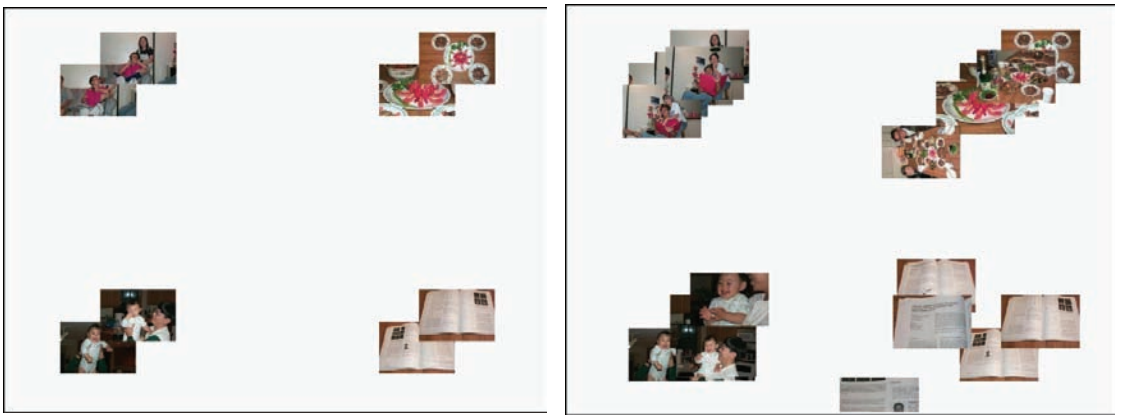
be recovered from pure 2D information. For comparison, Figure 10(c) shows the PCA Splat of the same images with their high-dimensional feature vectors weighted by a random α.

Figure 11 shows another example of user-guided layout. Assume that the user is describing her family story to a friend. In order not to disrupt the conversational flow, she only lays out a few photos from her personal photo collections and expects the computer to generate a similar and consistent layout for a larger set of images from the same collection. Figure 11(b) shows the computer-generated layout based on the learned feature weights from the configuration of Figure 11(a). The computer-generated layout is achieved using the α-estimation scheme and postlinear, for example, affine transform or nonlinear transformations. Only the 37 visual features (nine color moments (Stricker & Orengo 1995), 10 wavelet moments (Smith & Chang, 1994) and 18 water-filling features (Zhou et al., 1999)) were used for this PCA Splat. Clearly the computer-generated layout is similar to the user layout with the visually similar images positioned at the user-indicated locations. We should add that in this example no semantic features (keywords) were used, but it is clear that their addition would only enhance such a layout.

STATISTICAL ANALYSIS

Given the lack of sufficiently large (and willing) human subjects, we undertook a Monte Carlo approach to testing our user-modeling and estimation method. Monte Carlo simulation (Metropolis & Ulam, 1949) randomly generates values for uncertain variables over and over to simulate a model. Thereby simulating 1000 computer generated layouts (representing ground-truth values of α's), which were meant to emulate 1000 actual userlayouts or preferences. In each case, α estima-

Figure 11: Usermodeling for automatic layout. (a) a user-guided layout: (b) computer layout for larger set of photos (four classes and two photos from each class)



(a) *(b)*

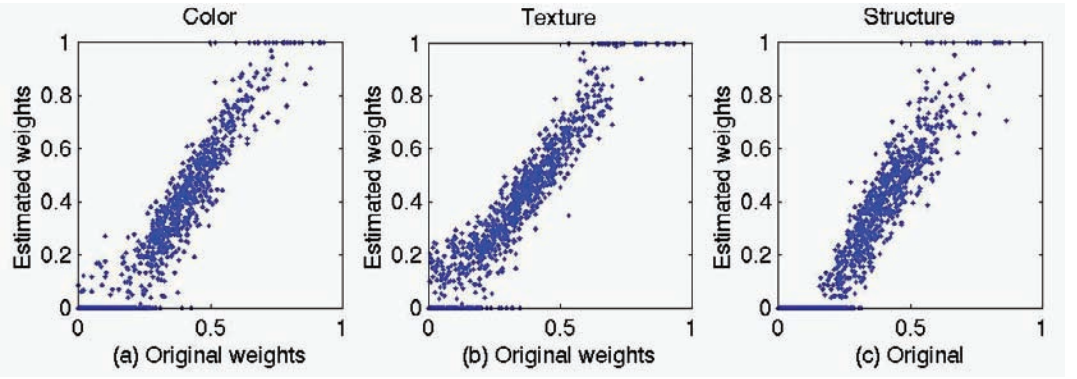
tion was performed to recover the original values as best as possible. Note that this recovery is only partially effective due to the information loss in projecting down to a 2D space. As a control, 1000 randomly generated feature weights were used to see how well they could match the user layouts (i.e., by chance alone).

Our primary test database consists of 142 images from the COREL database. It has 7 categories of car, bird, tiger, mountain, flower, church and airplane. Each class has about 20 images. Feature extraction based on color, texture and structure has been done off-line and prestored. Although we will be reporting on this test data set — due to its common use and familiarity to the CBIR community — we should emphasize that we have also successfully tested our methodology on larger and much more heterogeneous image libraries. (For example, real personal photo collections of 500+ images, including family, friends, vacations, etc.). Depending on the particular domain, one can obtain different degrees of performance, but one thing is for sure: for narrow application domain (for example, medical, logos, trademarks, etc.) it is quite easy to construct systems which work extremely well, by taking advantages of the limiting constraints in the imagery.

The following is the Monte Carlo procedure that was used for testing the significance and validity of user modeling with α estimation:

1. Randomly select *M* images from the database. Generate arbitrary (random) feature weights α in order to simulate a "user" layout.
2. Do a PCA Splat using this "ground truth" α.
3. From the resulting 2D layout, estimate α and denote the estimated α as .
4. Select a new distinct (nonoverlapping) set of *M* images from the database.
5. Do PCA Splats on the second set using the original α, the estimated , and a third random α' (as control).
6. Calculate the resulting stress in Equation (8), and layout deviation (2D position error) in Equation (9) for the original, estimated and random (control) values of α, , and α', respectively.
7. Repeat 1,000 times.

Figure 12. Scatter plot of estimation, estimated weights versus original weights



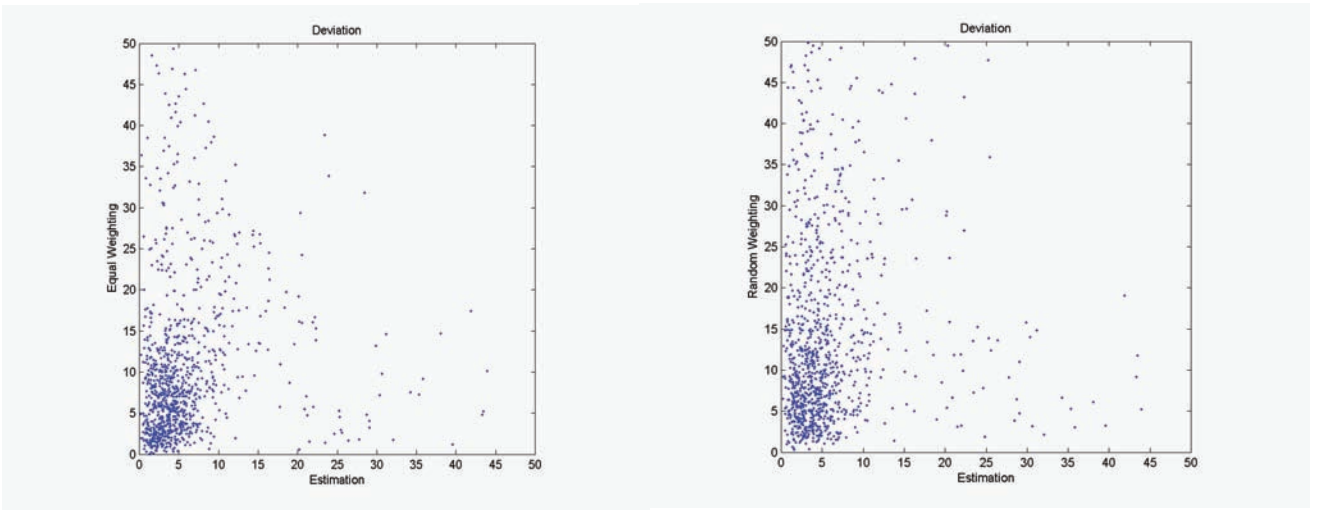
The scatter plot of α estimation is shown in Figure 12. Clearly there is a direct linear relationship between the original weights α and the estimated weights . Note that when the original weight is very small (<0.1) or very large (>0.9), the estimated weight is zero or one correspondingly. This means that when one particular feature weight is very large (or very small), the corresponding feature will become the most dominant (or least dominant) feature in the PCA, therefore the estimated weight for this feature will be either one or zero. This "saturation" phenomenon in Figure 12 is seen to occur more prominently for the case of structure (lower left of the rightmost panel) that is possibly more pronounced because of the structure feature vector being so (relatively) high dimensional. Additionally, structure features are not as well defined compared with color and texture (e.g., they have less discriminating power).

In terms of actual measures of stress and deviation we found that the α-estimation scheme yielded the smaller deviation 78.4% of the time and smaller stress 72.9%. The main reason these values are less than 100% is due to the nature of the Monte Carlo testing and the fact that working with low-dimensional (2D) spaces, random weights can be close to the original weights and hence can often generate similar "user" layouts (in this case apparently about 25% of the time).

We should add that an alternative "control" or null hypothesis to that of random weights α = is that of fixed equal weights. This "weighting" scheme corresponds to the assumption that there are to be no preferential biases in the subspace of the features, that they should all count equally in forming the final layout (or default PCA). But the fundamental premise behind the chapter is that there is a change or variable bias in the relative importance of the different features as manifested by different user layout and styles. In fact, if there was to be no bias in the weights (i.e., they were set equal) then there would be no usermodeling or adaptation necessary since there would always be just one type or "style" of layout (the one resulting from equal weights). In order to understand this question fully, we compare the results of random weights versus equal weights (compared to the estimation framework advocated).

In an identical set of experiments, replacing random weights for comparison layouts with equal weights $\alpha = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}^T$, we found a similar distribution of similarity scores. In particular, since the goal is obtaining accurate 2D layouts where positional accuracy is critical, we look at the resulting deviation in the case of both random weights and equal weight versus estimated weights. We carry out a large Monte Carlo experiments (10,000 trials) and Figure 13 shows the scatter plot of the

Figure 13. Scatter plot of deviation scores (a) equal weights (y-axis) versus estimation weights (x-axis) (b) random weights (y-axis) versus estimated weights (x-axis)



(a) Equal versus estimated

(b) Random versus estimated

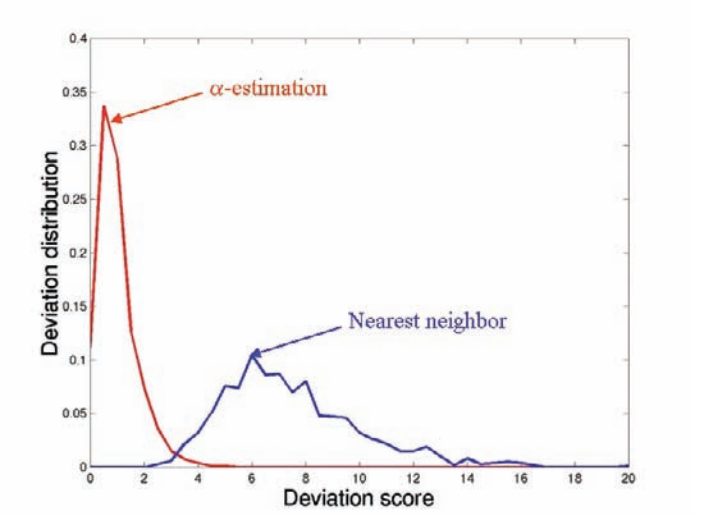
deviation scores. Points above the diagonal (not shown) indicate a deviation performance worse than that of weight estimation. As can be seen here, the test results are roughly comparable for equal and random weights.

In Figure 13, we noted that the α-estimation scheme yielded the smaller deviation 72.6% of the time compared to equal weights (as opposed to the 78.4% compared with random weights). We therefore note that the results and conclusions of these experiments are consistent despite the choice of equal or random controls, and ultimately direct α estimation of a "user" layout is best.

Finally we should note that all weighting schemes (random or not) define sensible or "coherent" layouts. The only difference is in the amount by which color, texture and structure is emphasized. Therefore even random weights generate "nice" or pleasing layouts, that is, random weights do not generate random layouts.

Another control other than random (or equal) weights is to compare the deviation of an α-estimation layout generator to a simple scheme which assigns each new image to the 2D location of its (un-weighted or equally weighted) 37-dimensional nearest neighbor (NN) from the set of images previously laid out by the "user." This control scheme essentially operates on the principle that new images should be positioned on screen at the same location as their nearest neighbors in the original 37-dimensional feature space (the default similarity measure in the absence of any prior bias) and thus essentially ignores the operating subspace defined by the "user" in a 2D layout. The NN placement scheme would place the test picture, despite their similarity score, directly on top of whichever image, currently on the "table", that it is closest to. To do otherwise, for example to place it slightly shifted away, and so forth, would simply imply the existence of a nondefault "smart" projection function which defeats the purpose of

Figure 14. Comparison of the distribution of a-estimation versus nearest-neighbor deviation scores



this "control." The point of this particular experiment is to compare our "smart" scheme with one which has no knowledge or preferential subspace weightings and see how this would subsequently map to (relative) position on the display. The idea behind that is that a dynamic user-centric display should adapt to varying levels of emphasis to color, texture, and structure.

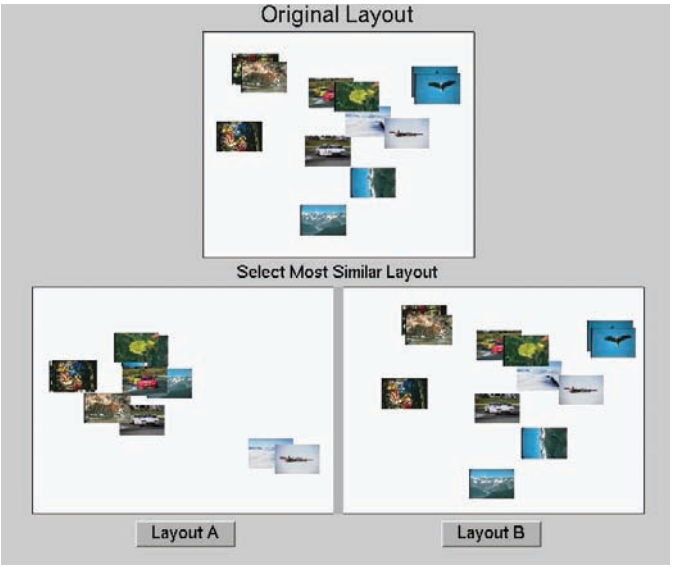
The distributions of the outcomes of this Monte Carlo simulation are shown in Figure 14 where we see that the layout deviation using α estimation (red: $\mu = 0.9691$, $\sigma = 0.7776$) was consistently lower — by almost an order of magnitude — than the nearest neighbor layout approach (blue: $\mu = 7.5921$ $\sigma = 2.6410$). We note that despite the noncoincident overlap of the distributions' tails in Figure 14, in every one of the 1,000 random trials the α-estimation deviation score was found to be smaller than that of nearest-neighbour (a key fact not visible in such a plot).

USER PREFERENCE STUDY

In addition to the computer-generated simulations, we have in fact conducted a preliminary user study that has also demonstrated the superior performance of α estimation over random feature weighting used as a control. The goal was to test whether the estimated feature weights would generate a better layout on a *new* but similar set of images than random weightings (used as control). The user interface is shown in Figure 15 where the top panel is a coherent layout generated by a random α on reference image set. From this layout, an estimate of α was computed and used to redo the layout. A layout generated according to random weights was also generated and used as a control. These two layouts were then displayed in the bottom panels with randomized (A vs. B) labels (in order to remove any bias in the presentation). The user's task was to select which layout (A or B) was more similar to the reference layout in the top panel.

In our experiment, six naïve users were instructed in the basic operation of the interface and given the following instructions: (1) both absolute and relative positions of images matter, (2) in general, similar images, like cars, tigers, and so forth, should cluster and (3) the relative positions of the clusters also matter. Each user performed 50 forced-choice tests with no time limits. Each test set of 50 contained redundant (randomly recurring) tests in order to test the user's consistency.

Figure 15. α-estimation-matters user test interface



We specifically aimed at not "priming" the subjects with very detailed instructions (such as, "It's not valid to match a red car and a red flower because they are both red."). In fact, the "naïve" test subjects were told nothing at all about the three feature types (color, texture, structure), the associated α or obviously the estimation technique. In this regard, the paucity of the instructions was entirely intentional: whatever mental grouping that seemed valid to them was the key. In fact, this very same flexible association of the user is what was specifically tested for in the *consistency* part of the study.

Table 1 shows the results of this user study. The average preference indicated for the α-estimation-based layout was found to be 96% and an average consistency rate of a user was 97%. We note that the α-estimation method of generating a layout in a similar "style" to the reference was consistently favored by the users. A similar experimental study has shown this to also be true even if the test layouts consist of *different* images than those used in the reference layout (i.e., similar but not identical images from the same categories or classes).

DISCUSSIONS AND FUTURE WORK

We have designed our system with general CBIR in mind but more specifically for personalized photo collections. An optimized content-based visualization technique is proposed to generate a 2D display of the retrieved images for content-based image retrieval. We believe that both the computational results and the pilot user study support our claims of a more perceptually intuitive and informative visualization engine that not only provides a better understanding of query retrievals but also aids in forming new queries.

The proposed content-based visualization method can be easily applied to project the images from high-dimensional feature space to a 3D space for more advanced visualization and navigation. Features can be multimodal, expressing individual visual features, for example, color

Table 1. Results of user-preference study

	P reference for estimates	Preference for random weights	User's consistency rate
User 1	90%	10%	100%
User 2	98%	2 %	90%
User 3	98%	2 %	90%
User 4	95%	5 %	100%
User 5	98%	2 %	100%
User 6	98%	2 %	100%
Average	**96%**	**4 %**	**97%**

alone, audio features and semantic features, for example, keywords, or any combination of the above. The proposed layout optimization technique is also quite general and can be applied to avoid overlapping of any type of images, windows, frames or boxes.

The PDH project is at its initial stage. We have just begun our work in both the user interface design and photo visualization and layout algorithms. The final visualization and retrieval interface can be displayed on a computer screen, large panel projection screens, or, for example, on embedded tabletop devices (Shen et al., 2001; Shen et al., 2003) designed specifically for purposes of storytelling or multiperson collaborative exploration of large image libraries.

Many interesting questions still remain as our future research in the area of content-based information visualization and retrieval. The next task is to carry out an extended user-modeling study by having our system learn the feature weights from various sample layouts provided by the user. We have already developed a framework to incorporate visual features with semantic labels for both retrieval and layout.

Another challenging area is automatic "summarization" and display of large image collections. Since summarization is implicitly defined by user preference, a estimation for user modeling will play a key role in this and other high-level tasks where context is defined by the user.

Finally, incorporation of relevance feedback for content-based image retrieval based on the visualization of the optimized PCA Splat seems very intuitive and is currently being explored. By manually grouping the relevant images together at each relevance feedback step, a dynamic user-modeling technique will be proposed.

ACKNOWLEDGMENT

This work was supported in part by Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, and National Science Foundation Grant EIA 99-75019.

REFERENCES

Balabanovic, M., Chu, L., & Wolff, G. (2000). Storytelling with digital photographs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, The Hague, The Netherlands (pp. 564-571).

Belkin, M., & Niyogi, P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, *15*(6),1373-1396.

Brand, M. (2003). *Charting a manifold*. Mitsubishi Electric Research Laboratories (MERL), TR2003-13.

Dietz, P., & Leigh, D. (2001). DiamondTouch: A multi-user touch technology. *The Proceedings of the 14th ACM Symposium on User Interface Software and Technology*, Orlando, Florida (pp. 219-226).

Horoike, A., & Musha, Y. (2000). Similarity-based image retrieval system with 3D visualization. *Proceedings of IEEE International Conference on Multimedia and Expo*, New York, New York (Vol. 2, pp. 769-772).

Jolliffe, I. T. (1996). *Principal component analysis.* New-York: Springer-Verlag.

Kang, H., & Shneiderman, B. (2000). Visualization methods for personal photo collections: Browsing and searching in the photofinder. *Proceedings of IEEE International Conference on Multimedia and Expo*, New York, New York.

Metropolis, N. & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association, 44*(247), 335-341.

Moghaddam, B., Tian, Q., & Huang, T. S. (2001). Spatial visualization for content-based image retrieval. *Proceedings of IEEE International Conference on Multimedia and Expo*, Tokyo, Japan.

Moghaddam, B., Tian, Q., Lesh, N., Shen, C., & Huang, T.S. (2002). PDH: A human-centric interface for image libraries. *Proceedings of IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland (Vol. 1, pp. 901-904).

Moghaddam, B., Tian, Q., Lesh, N., Shen, C., & Huang, T.S. (2004). Visualization and user-modeling for browsing personal photo libraries. *International Journal of Computer Vision, Special Issue on Content-Based Image Retrieval, 56*(1-2), 109-130.

Nakazato, M., & Huang, T.S. (2001). 3D MARS: Immersive virtual reality for content-based image retrieval. *Proceedings of IEEE International Conference on Multimedia and Expo*, Tokyo, Japan.

Popescu, M., & Gader, P. (1998). Image content retrieval from image databases using feature integration by choquet integral. *Proceeding of SPIE Conference on Storage and Retrieval for Image and Video Databases VII*, San Jose, California.

Roweis, S., & Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*(5500), 2323-2326.

Rubner, Y. (1999). *Perceptual metrics for image database navigation*. Doctoral dissertation, Stanford University.

Santini, S., Gupta, A., & Jain, R. (2001). Emergent semantics through interaction in image databases. *IEEE Transactions on Knowledge and Data Engineering, 13*(3), 337-351.

Santini, S., & Jain, R. (1999). Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(9), 871-883.

Santini, S., & Jain, R., (2000, July-December). Integrated browsing and querying for image databases. *IEEE Multimedia Magazine*, 26-39.

Shen, C., Lesh, N., & Vernier, F. (2003). Personal digital historian: Story sharing around the table. *ACM Interactions*, March/April (also MERL TR2003-04).

Shen, C., Lesh, N., Moghaddam, B., Beardsley, P., & Bardsley, R. (2001). Personal digital historian: User interface design. *Proceedings of Extended Abstract of SIGCHI Conference on Human Factors in Computing Systems*, Seattle, Washington (pp. 29-30).

Shen, C., Lesh, N., Vernier, F., Forlines, C., & Frost, J. (2002). *Sharing and building digital group histories.* ACM Conference on Computer Supported Cooperative Work.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349-1380.

Smith, J. R., & Chang, S. F. (1994). Transform features for texture classification and discrimination in large image database. *Proceedings of IEEE International Conference on Image Processing*, Austin, TX.

Squire, D. M., Müller, H., & Müller, W. (1999). Improving response time by search pruning in a content-based image retrieval system using inverted file techniques. *Proceedings of IEEE Workshop On Content-Based Access of Image and Video Libraries (CBAIVL),* Fort Collins, CO.

Stricker, M., & Orengo, M. (1995). Similarity of color images. *Proceedings of. SPIE Storage and Retrieval for Image and Video Databases*, San Diego, CA.

Swets, D., & Weng, J. (1999). Hierarchical discriminant analysis for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(5), 396-401.

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, 290*, 2319-2323.

Tian, Q., Moghaddam, B., & Huang, T. S. (2001). Display optimization for image browsing. *The Second International Workshop on Multimedia Databases and Image Communications*, Amalfi, Italy (pp. 167-173).

Tian, Q., Moghaddam, B., & Huang, T. S. (2002). Visualization, estimation and user-modeling for interactive browsing of image libraries. *International Conference on Image and Video Retrieval*, London (pp. 7-16).

Torgeson, W. S. (1998). *Theory and methods of scaling.* New York: John Wiley & Sons.

Vernier, F., Lesh, N., & Shen, C. (2002). Visualization techniques for circular tabletop interface. *Proceedings of Advanced Visual Interfaces (AVI),* Trento, Italy (pp. 257-266).

Zhou, S. X., Rui, Y., & Huang, T. S. (1999). Water-filling algorithm: A novel way for image feature extraction based on edge maps. *Proceedings of IEEE International Conference on Image Processing*, Kobe, Japan.

Zwillinger, D. (Ed.) (1995). *Affine transformations, §4.3.2 in CRC standard mathematical tables and formulae* (pp. 265-266). Boca Raton, FL: CRC Press.

APPENDIX A

Let

$$\mathbf{P}_i = p^{(i)}(x,y) = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

and

$$\hat{\mathbf{P}}_i = \hat{p}^{(i)}(x,y) = \begin{pmatrix} \hat{x}_i \\ \hat{y}_i \end{pmatrix},$$

and Equation (15) is rewritten as

$$J = \sum_{i=1}^{N} \| \mathbf{P}_i - \hat{\mathbf{P}}_i \|^2 \qquad \text{(A.1)}$$

Let $\mathrm{X_i}$ be the column feature vector of the i^{th} image, where

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{X}_c^{(i)} \\ \mathbf{X}_t^{(i)} \\ \mathbf{X}_s^{(i)} \end{bmatrix},$$

$i = 1, \cdots, N$. $\mathbf{X}_c^{(i)}$, $\mathbf{X}_t^{(i)}$ and $\mathbf{X}_s^{(i)}$ are the corresponding color, texture and structure feature vector of the i^{th} image and their lengths are L_c, L_t and L_s, respectively. Let

$$\mathbf{X}'_{\mathrm{i}} = \begin{pmatrix} \alpha_{\mathrm{c}}\mathbf{X}_{\mathrm{c}}^{(\mathrm{i})} \\ \alpha_{\mathrm{t}}\mathbf{X}_{\mathrm{t}}^{(\mathrm{i})} \\ \alpha_{\mathrm{s}}\mathbf{X}_{\mathrm{s}}^{(\mathrm{i})} \end{pmatrix}$$

be the weighted high-dimensional feature vector. These weights α_c, α_t, α_s are constrained such as they always sum to 1.

$\hat{\mathbf{P}}_i$ is estimated by linearly projecting the weighted high-dimensional features to 2D. Let $\mathrm{X} = [\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_N]$, it is an $L \times N$ matrix, where $L = L_c + L_t + L_s$. $\hat{\mathbf{P}}_i$ is estimated by

$$\hat{\mathbf{P}}_{\mathrm{i}} = U^T(\mathbf{X}'_{\mathrm{i}} - \mathbf{X}_{\mathrm{m}})\, i = 1, \cdots, N \qquad \text{(A.2)}$$

where U is a $L \times 2$ projection matrix, X_{m} is an $L \times 1$ mean column vector of X'_{i}, $i = 1, ..., N$. Substitute $\hat{\mathbf{P}}_i$ by Equation (A.2) into Equation (A.1), the problem is therefore one of seeking the optimal feature weights α, projection matrix U, and column vector X_m such as J in Equation (A.3) is minimized, given X_{i}, P_{i}, $i = 1, ..., N$.

$$J = \sum_{i=1}^{N} \| U^T(\mathbf{X}'_{\mathrm{i}} - \mathbf{X}_m) - \mathbf{P}_{\mathrm{i}} \|^2 \qquad \text{(A.3)}$$

In practice, it is almost impossible to estimate optimal α, U and X_{m} simultaneously based on the limited available data X_{i}, P_{i}, $i = 1, ..., N$. We thus make some modifications. Instead of estimating α, U and X_m simultaneously, we modified the estimation process to be a two-step *re-estimation* procedure. We first estimate the projection matrix U and column vector X_m, and then estimate feature weight vector α based on the computed U and X_m, and iterate until convergence.

Let $U^{(0)}$ be the eigenvectors corresponding to the largest two eigenvalues of the covariance matrix of $\mathbf{X}$, where $\mathrm{X} = [\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_N]$, is the mean vector of $\mathbf{X}$.

We have

$$\mathbf{P}_{\mathrm{i}}^{(0)} = (U^{(0)})^T(\mathbf{X}_{\mathrm{i}} - \mathbf{X}_{\mathrm{m}}^{(0)}) \qquad \text{(A. 4)}$$

$\mathbf{P}_{\mathrm{i}}^{(0)}$ is the projected 2D coordinates of the unweighted high-dimensional feature vector of the i^{th} image. Ideally its target location is $\mathbf{P}_{\mathrm{i}}$. To consider the alignment correction, a rigid transform (Zwillinger, 1995) is applied.

$$\hat{\mathbf{P}}_{\mathrm{i}}^{(0)} = A \cdot \mathbf{P}_{\mathrm{i}}^{(0)} + T \qquad \text{(A. 5)}$$

where A is a 2×2 matrix and T is a 2×1 vector. A, T are obtained by minimizing the L_2-norm of $\mathbf{P}_{\mathrm{i}} - \hat{\mathbf{P}}_{\mathrm{i}}^{(0)}$.

Therefore J in Equation (A. 3) is modified to

$$J = \sum_{i=1}^{N} \| AU^{(0)}(\mathbf{X}'_i - \mathbf{X}_{\mathrm{m}}^{(0)}) - (\mathbf{P}_{\mathrm{i}} - T) \|^2 \qquad \text{(A. 6)}$$

Let $U = UA^{(0)}$, $\mathbf{X}_m = \mathbf{X}_{\mathrm{m}}^{(0)}$ and $\mathbf{P}_{\mathrm{i}} = \mathbf{P}_{\mathrm{i}} - T$, we still have the form of Equation (A. 3).

Let us rewrite

$$U^T = \begin{bmatrix} U_{11}, \cdots U_{1(L_c+L_t+L_s)} \\ U_{21}, \cdots, U_{2(L_c+L_t+L_s)} \end{bmatrix}$$

After some simplifications on Equation (A. 3), we have

$$J = \sum_{i=1}^{N} \| \alpha_c A_i + \alpha_t B_i + \alpha_s C_i - D_i \|^2 \qquad \text{(A. 7)}$$

where

$$A_i = \begin{bmatrix} \sum_{k=1}^{L_c} U_{1k}\mathbf{X}_c^{(k)}(i) \\ \sum_{k=1}^{L_c} U_{2k}\mathbf{X}_c^{(k)}(i) \end{bmatrix} \qquad B_i = \begin{bmatrix} \sum_{k=1}^{L_t} U_{1(k+L_c)}\mathbf{X}_t^{(k)}(i) \\ \sum_{k=1}^{L_t} U_{2(k+L_t)}\mathbf{X}_t^{(k)}(i) \end{bmatrix}$$

$$C_i = \begin{bmatrix} \sum_{k=1}^{L_s} U_{1(k+L_c+L_t)}\mathbf{X}_s^{(k)}(i) \\ \sum_{k=1}^{L_s} U_{2(k+L_c+L_t)}\mathbf{X}_s^{(k)}(i) \end{bmatrix}$$

and $D_i = U^T\mathbf{X}_m + \mathbf{P}_i$,

A_i, B_i, C_i and D_i are the 2×1feature vectors, respectively.

To minimize J, we take the partial derivatives of J relative to α_c, α_t, α_s and set them to zero, respectively.

$$\begin{aligned} \frac{\partial J}{\partial \alpha_c} &= 0 \\ \frac{\partial J}{\partial \alpha_t} &= 0 \\ \frac{\partial J}{\partial \alpha_s} &= 0 \end{aligned} \quad \text{(A. 8)}$$

We thus have:

$$E \cdot \boldsymbol{\alpha} = f \quad \text{(A. 9)}$$

where

$$E = \begin{bmatrix} \sum_{i=1}^{N} A_i^T A_i & \sum_{i=1}^{N} A_i^T B_i & \sum_{i=1}^{N} A_i^T C_i \\ \sum_{i=1}^{N} B_i^T A_i & \sum_{i=1}^{N} B_i^T B_i & \sum_{i=1}^{N} B_i^T C_i \\ \sum_{i=1}^{N} C_i^T A_i & \sum_{i=1}^{N} C_i^T B_i & \sum_{i=1}^{N} C_i^T C_i \end{bmatrix},$$

$$f = \begin{bmatrix} \sum_{i=1}^{N} A_i^T D_i \\ \sum_{i=1}^{N} B_i^T D_i \\ \sum_{i=1}^{N} C_i^T D_i \end{bmatrix}$$

and α is obtained by solving the linear Equation (A.9).

This work was previously published in Managing Multimedia Semantics, edited by U. Srinivasan and S. Nepal, pp. 193-222, copyright 2005 by IRM Press (an imprint of IGI Global).

Chapter 7.10
Digital Signature–Based Image Authentication

Der-Chyuan Lou
National Defense University, Taiwan

Jiang-Lung Liu
National Defense University, Taiwan

Chang-Tsun Li
University of Warwick, UK

ABSTRACT

This chapter is intended to disseminate the concept of digital signature-based image authentication. Capabilities of digital signature-based image authentication and its superiority over watermarking-based approaches are described first. Subsequently, general models of this technique — strict authentication and non-strict authentication are introduced. Specific schemes of the two general models are also reviewed and compared. Finally, based on the review, design issues faced by the researchers and developers are outlined.

INTRODUCTION

In the past decades, the technological advances of international communication networks have facilitated efficient digital image exchanges. However, the availability of versatile digital signal/image processing tools has also made image duplication trivial and manipulations discernable for the human visual system (HVS). Therefore, image authentication and integrity verification have become a popular research area in recent years. Generally, image authentication is projected as a procedure of guaranteeing that the image content has not been altered, or at least that the visual (or semantic) characteristics of the image are maintained after incidental manipulations such as JPEG compression. In other words, one of the objectives of image authentication is to verify the integrity of the image. For many applications such as medical archiving, news reporting and political events, the capability of detecting manipulations of digital images is often required. Another need for image authentication arises from the requirement of checking the identity of the image sender. In the scenario that a buyer wants to purchase and receive an image over the networks, the buyer may obtain the image via e-mails or from the

Internet-attached servers that may give a malicious third party the opportunities to intercept and manipulate the original image. So the buyer needs to assure that the received image is indeed the original image sent by the seller. This requirement is referred to as the *legitimacy* requirement in this chapter.

To address both the integrity and legitimacy issues, a wide variety of techniques have been proposed for image authentication recently. Depending on the ways chosen to convey the authentication data, these techniques can be roughly divided into two categories: *labeling-based techniques* (e.g., the method proposed by Friedman, 1993) and *watermarking-based techniques* (e.g., the method proposed by Walton, 1995). The main difference between these two categories of techniques is that labeling-based techniques create the authentication data in a separate file while watermarking-based authentication can be accomplished without the overhead of a separate file. However, compared to watermarking-based techniques, labeling-based techniques potentially have the following advantages.

- They can detect the change of every single bit of the image data if strict integrity has to be assured.
- The image authentication can be performed in a secure and robust way in public domain (e.g., the Internet).
- The data hiding capacity of labeling-based techniques is higher than that of watermarking.

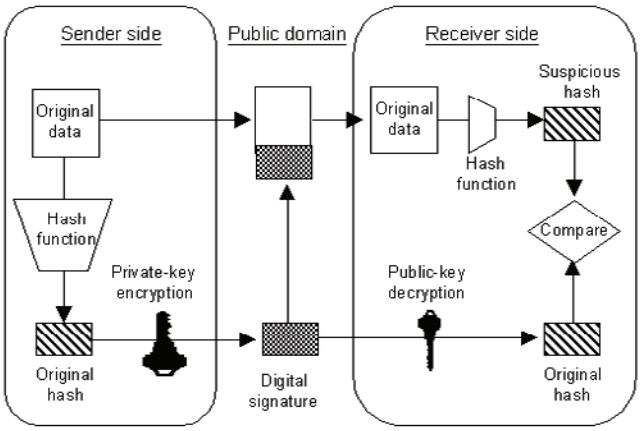
Given its advantages on watermarking-based techniques, we will focus on labeling-based authentication techniques.

In labeling-based techniques, the authentication information is conveyed in a separate file called *label*. A label is additional information associated with the image content and can be used to identify the image. In order to associate the label content with the image content, two different ways can be employed and are stated as follows.

- The first methodology uses the functions commonly adopted in message authentication schemes to generate the authentication data. The authentication data are then encrypted with secret keys or private keys depending on what cryptographic authentication protocol is employed. When applying to two different bit-streams (i.e., different authentication data), these functions can produce two different bit sequences, in such a way that the change of every single bit of authentication data can be detected. In this chapter, image authentication schemes of this class are referred to as *strict authentication.*
- The second methodology uses some special-purpose functions to extract essential image characteristics (or *features*) and encrypt them with senders' private keys (Li, Lou & Chen, 2000; Li, Lou & Liu, 2003). This procedure is the same as the digital signature protocol except that the features must be designed to compromise with some specific image processing techniques such as JPEG compression (Wallace, 1991). In this chapter, image authentication techniques of this class are referred to as *non-strict authentication.*

The strict authentication approaches should be used when strict image integrity is required and no modification is allowed. The functions used to produce such authentication data (or *authenticators*) can be grouped into three classes: *message encryption, message authentication code* (MAC), and *hash function* (Stallings, 2002). For *message encryption,* the original message is encrypted. The encrypted result (or cipher-text) of the entire message serves as its authenticator. To authenticate the content of an image, both the sender and receiver share the same secret key.

Figure 1. Process of digital signature



Message authentication code is a fixed-length value (authenticator) that is generated by a public function with a secret key. The sender and receiver also share the same secret key that is used to generate the authenticator. A *hash function* is a public function that maps a message of any length to a fixed-length hash value that serves as the authenticator. Because there is no secret key adopted in creating an authenticator, the hash functions have to be included in the procedure of digital signature for the electronic exchange of message. The details of how to perform those labeling-based authentication schemes and how to obtain the authentication data are described in the second section.

The non-strict authentication approaches must be chosen when some forms of image modifications (e.g., JPEG lossy compression) are permitted, while malicious manipulation (e.g., objects' deletion and modification) must be detected. This task can be accomplished by extracting features that are invariant to predefined image modifications. Most of the proposed techniques in the literature adopted the same authentication procedure as that performed in digital signature to resolve the legitimacy problem, and exploited invariant features of images to resolve the non-strict authentication. These techniques are often regarded as digital signature-based techniques and will be further discussed in the rest of this chapter. To make the chapter self-contained, some labeling-based techniques that do not follow the standard digital-signature procedures are also introduced in this chapter.

This chapter is organized as follows. Following the introduction in the first section, the second section presents some generic models including strict and non-strict ones for digital signature-based image authentication. This is followed by a section discussing various techniques for image authentication. Next, the chapter addresses the challenges for designing secure digital signature-based image authentication methods. The final section concludes this chapter.

GENERIC MODELS

The digital signature-based image authentication is based on the concept of digital signature, which

is derived from a cryptographic technique called public-key cryptosystem (Diffie & Hellman, 1976; Rivest, Shamir & Adleman, 1978). Figure 1 shows the basic model of digital signature. The sender first uses a hash function, such as MD5 (Rivest, 1992), to hash the content of the original data (or *plaintext*) to a small file (called *digest*). Then the digest is encrypted with the sender's private key. The encrypted digest can form a unique "signature" because only the sender has the knowledge of the private key. The signature is then sent to the receiver along with the original information. The receiver can use the sender's public key to decrypt the signature, and obtain the original digest. Of course, the received information can be hashed by using the same hash function in the sender side. If the decrypted digest matches the newly created digest, the legitimacy and the integrity of the message are therefore authenticated.

There are two points worth noting in the process of digital signature. First, the plaintext is not limited to text file. In fact, any types of digital data, such as digitized audio data, can be the original data. Therefore, the original data in Figure 1 can be replaced with a digital image, and the process of digital signature can then be used to verify the legitimacy and integrity of the image. The concept of trustworthy digital camera (Friedman, 1993) for image authentication is based on this idea. In this chapter, this type of image authentication is referred to as *digital signature-based image authentication*. Second, the hash function is a mathematical digest function. If a single bit of the original image is changed, it may result in a different hash output. Therefore, the strict integrity of the image can be verified, and this is called strict authentication in this chapter. The framework of strict authentication is described in the following subsection.

Strict Authentication

Figure 2 shows the main elements and their interactions in a generic digital signature-based model for image authentication. Assume that the sender wants to send an image I to the receiver, and the legitimate receiver needs to assure the legitimacy and integrity of I. The image I is first hashed to a small file h. Accordingly:

$$h = \mathrm{H}(I), \tag{1}$$

where H(·) denotes hash operator. The hashed result h is then encrypted (signed) with the sender's private key K_R to generate the signature:

$$S = \mathrm{E}_{K_R}(h), \tag{2}$$

where E(·) denotes the public-key encryption operator. The digital signature S is then attached to the original image to form a composite message:

$$M = I \parallel S, \tag{3}$$

where "||" denotes concatenation operator.

If the legitimacy and integrity of the received image I' needs to be verified, the receiver first separates the suspicious image I' from the composite message, and hashes it to obtain the new hashed result, that is:

$$h' = \mathrm{H}(I'). \tag{4}$$

The attached signature is decrypted with the sender's public-key K_p to obtain the possible original hash code:

$$\hat{h} = \mathrm{D}_{Kp}(\hat{S}), \tag{5}$$

where D(·) denotes the public-key decryption operator. Note that we use $\hat{S}$ and $\hat{h}$ respectively to represent the received signature and its hash result because the received signature may be a forged one. The legitimacy and integrity can be confirmed by comparing the newly created hash h' and the possible original hash $\hat{h}$. If they match with each other, we can claim that the received image I' is authentic.

Figure 2. Process of digital signature-based strict authentication

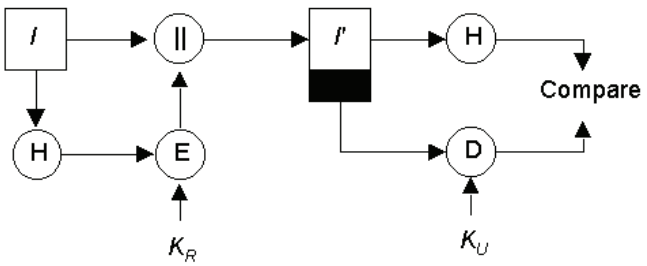
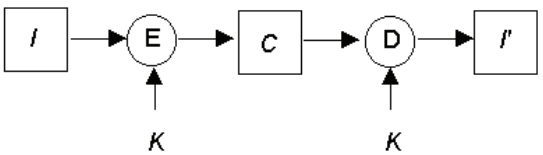


Figure 3. Process of encryption function-based strict authentication

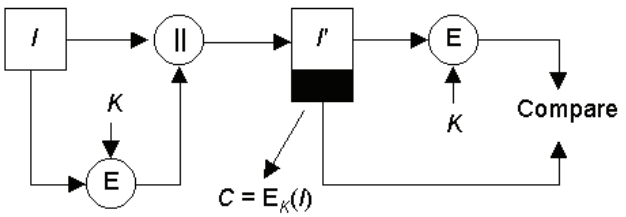


The above framework can be employed to make certain the strict integrity of an image because of the characteristics of the hash functions. In the process of digital signature, one can easily create the hash of an image, but it is difficult to reengineer a hash to obtain the original image. This can be also referred to "*one-way*" property. Therefore, the hash functions used in digital signature are also called *one-way hash* functions. MD5 and SHA (NIST FIPS PUB, 1993) are two good examples of one-way hash functions. Besides one-way hash functions, there are other authentication functions that can be utilized to perform the strict authentication. Those authentication functions can be classified into two broad categories: conventional encryption functions and message authentication code (MAC) functions.

Figure 3 illustrates the basic authentication framework for using conventional encryption functions. An image, *I*, transmitted from the sender to the receiver, is encrypted using a secret key *K* that was shared by both sides. If the decrypted image *I'* is meaningful, then the image is authentic. This is because only the legitimate sender has the shared secret key. Although this is a very straightforward method for strict image authentication, it also provides opponents opportunities to forge a meaningful image. For example, if an opponent has the pair of (*I, C*), he/she can forge an intelligible image *I'* by the *cutting and pasting* method (Li, Lou & Liu, 2003). One solution to this problem is to use the message authentication code (MAC).

Figure 4 demonstrates the basic model of MAC-based strict authentication. The MAC is a cryptographic checksum that is first generated with a shared secret key before the transmission of the original image *I*. The MAC is then transmitted to the receiver along with the original image. In order to assure the integrity, the receiver conducts the same calculation on the received image *I'* using the same secret key to generate a new MAC. If the received MAC matches the calculated MAC, then the integrity of the received image is

Figure 4. Process of MAC-based strict authentication



verified. This is because if an attacker alters the original image without changing the MAC, then the newly calculated MAC will still differ from the received MAC.

The MAC function is similar to the encryption one. One difference is that the MAC algorithm does not need to be reversible. Nevertheless, the decryption formula must be reversible. It results from the mathematical properties of the authentication function. It is less vulnerable to be broken than the encryption function. Although MAC-based strict authentication can detect the fake image created by an attacker, it cannot avoid "legitimate" forgery. This is because both the sender and the receiver share the same secret key. Therefore, the receiver can create a fake image with the shared secret key, and claim that this created image is received from the legitimate sender.

With the existing problems of encryption and MAC functions, the digital signature-based method seems a better way to perform strict authentication. Following the increasing applications that can tolerate one or more content-preserving manipulations, non-strict authentication becomes more and more important nowadays.

Non-Strict Authentication

Figure 5 shows the process of non-strict authentication. As we can see, the procedure of non-strict authentication is similar to that of strict authentication except that the function here used to digest the image is a special-design feature extraction function f_C.

Assume that the sender wants to deliver an image I to the receiver. A feature extraction function f_C is used to extract the image feature and to encode it to a small feature code:

$$C = f_C(I), \tag{6}$$

where $f_C(\cdot)$ denotes feature extraction and coding operator. The extracted feature code has three significant properties. First, the size of extracted feature code is relatively small compared to the size of the original image. Second, it preserves the characteristics of the original image. Third, it can tolerate incidental modifications of the original image. The feature code C is then encrypted (signed) with the sender's private key K_R to generate the signature:

$$S = \mathrm{E}_{K_R}(C). \tag{7}$$

The digital signature S is then attached to the original image to form a composite message:

$$M = I \parallel S. \tag{8}$$

Then the composite message M is forwarded to the receiver. The original image may be lossy compressed, decompressed, or tampered during transmission. Therefore, the received composite message may include a corrupted image I'. The original I may be compressed prior to the concatenation operation. If a lossy compression strategy

is adopted, the original image *I* in the composite message can be considered as a corrupted one.

In order to verify the legitimacy and integrity of the received image *I'*, the receiver first separates the corrupted image *I'* from the composite message, and generates a feature code *C'* by using the same feature extraction function in the sender side, that is:

$$C' = f_C(I'). \quad (9)$$

The attached signature is decrypted with the sender's public-key K_U to obtain the original feature code:

$$\hat{C} = \mathrm{D}_{K_U}(\hat{S}). \quad (10)$$

Note that we use $\hat{S}$ and $\hat{C}$ to represent the received signature and feature code here because the signature may be forged.

The legitimacy and integrity can be verified by comparing the newly generated feature *C'* and the received feature code $\hat{C}$. To differentiate the errors caused by authorized modifications from the errors of malevolent manipulations, let *d*(*C*, *C'*) be the measurement of similarity between the extracted features and the original. Let *T* denote a tolerable threshold value for examining the values of *d*(*C*, *C'*) (e.g., it can be obtained by performing a maximum compression to an image). The received image may be considered authentic if the condition < *T* is met.

Defining a suitable function to generate a feature code that satisfies the requirements for non-strict authentication is another issue. Ideally, employing a feature code should be able to detect content-changing modifications and tolerate content-preserving modifications. The content-changing modifications may include cropping, object addition, deletion, and modification, and so forth, while the content-preserving modifications may include lossy compression, format conversion and contrast *enhancing*, etc.

It is difficult to devise a feature code that is sensitive to all the content-changing modifications, while it remains insensitive to all the content-preserving modifications. A practical approach to design a feature extraction function would be based on the manipulation methods (e.g., JPEG lossy compression). As we will see in the next section, most of the proposed non-strict authentication techniques are based on this idea.

STATE OF THE ART

In this section, several existing digital signature-based image authentication schemes are detailed. Specifically, works related strict authentication is described in the first subsection and non-strict ones in the second subsection. Note that the intention of this section is to describe the methodology of the techniques. Some related problems about these techniques will be further discussed in the fourth section, in which some issues of designing practical schemes of digital signature-based image authentication are also discussed.

Strict Authentication

Friedman (1993) associated the idea of digital signature with digital camera, and proposed a "trustworthy digital camera," which is illustrated as Figure 6. The proposed digital camera uses a digital sensor instead of film, and delivers the image directly in a computer-compatible format. A secure microprocessor is assumed to be built in the digital camera and be programmed with the private key at the factory for the encryption of the digital signature. The public key necessary for later authentication appears on the camera body as well as the image's border. Once the digital camera captures the objective image, it produces two output files. One is an all-digital industry-standard file format representing the captured image; the other is an encrypted digital signature generated by applying the camera's unique private key (embedded in the camera's secure microprocessor) to a hash of the captured image

Figure 6. Idea of the trustworthy digital camera

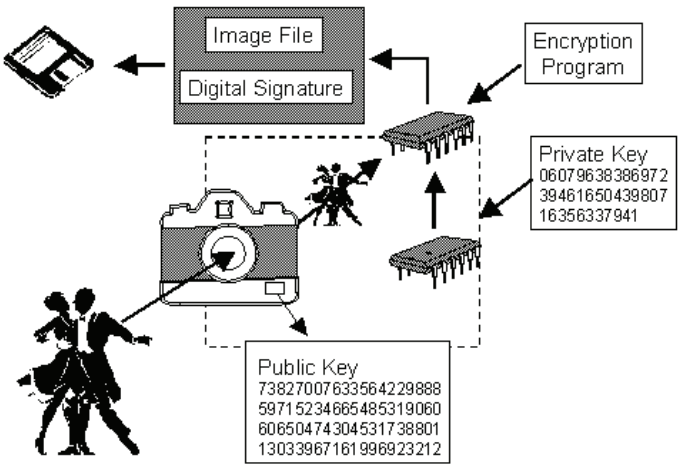
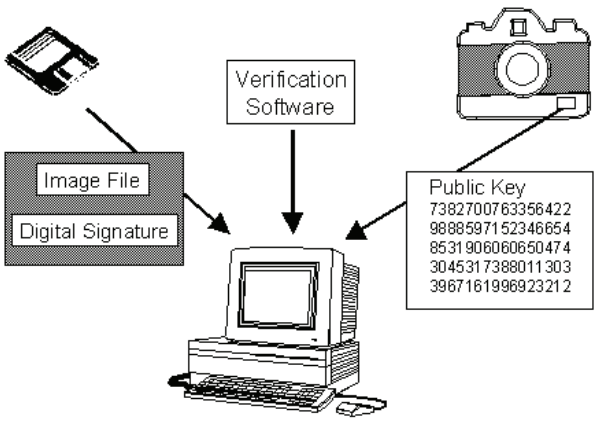


Figure 7. Verification process of Friedman's idea



file, a procedure described in the second section. The digital image file and the digital signature can later be distributed freely and safely.

The verification process of Friedman's idea is illustrated in Figure 7. The image authentication can be accomplished with the assistance of the public domain verification software. To authenticate a digital image file, the digital image, its accompanying digital signature file, and the public key are needed by the verification software running on a standard computer platform. The program then calculates the hash of the input image, and uses the public key to decode the digital signature to reveal the original hash. If these two hash values match, the image is considered to be authentic. If these two hash values are different, the integrity of this image is questionable.

It should be noted that the hash values produced by using the cryptographic algorithm such as MD5 will not match if a single bit of the image file is changed. This is the characteristic of the strict authentication, but it may not be suitable for authenticating images that undergo lossy compression. In this case, the strict authentication code (hash values) should be generated in a non-strict way. Non-strict authentication schemes have been proposed for developing such algorithms.

Non-Strict Authentication

Instead of using a strict authentication code, Schneider and Chang (1996) used content-based data as the authentication code. Specifically, the content-based data can be considered to be the image feature. As the image feature is invariant for some content-preserving transformation, the original image can also be authenticated although it may be manipulated by some allowable image transformations. The edge information, DCT coefficients, color, and intensity histograms are regarded as potentially invariant features. In Schneider and Chang's method, the intensity histogram is employed as the invariant feature in the implementation of the content-based image authentication scheme. To be effective, the image is divided into blocks of variable sizes and the intensity histogram of each block is computed separately and is used as the authentication code.

To tolerate incidental modifications, the Euclidean distance between intensity histograms was used as a measure of the content of the image. It is reported that the lossy compression ratio that could be applied to the image without producing a false positive is limited to 4:1 at most. Schneider and Chang also pointed out that using a reduced distance function can increase the maximum permissible compression ratio. It is found that the alarm was not triggered even at a high compression ratio up to 14:1 if the block average intensity is used for detecting image content manipulation. Several works have been proposed in the literature based on this idea. They will be introduced in the rest of this subsection.

Feature-Based Methods

The major purpose of using the image digest (hash values) as the signature is to speed up the signing procedure. It will violate the principle of the digital signature if large-size image features were adopted in the authentication scheme. Bhattacharjee and Kutter (1998) proposed another algorithm to extract a smaller size feature of an image. Their feature extraction algorithm is based on the so-called *scale interaction model*. Instead of using Gabor wavelets, they adopted Mexican-Hat wavelets as the filter for detecting the feature points. The algorithm for detecting feature-points is depicted as follows.

- Define the feature-detection function, $P_{ij}(\cdot)$ as:

$$P_{ij}(\vec{x}) = | M_i(\vec{x}) - \gamma \cdot M_j(\vec{x}) | \qquad (11)$$

  where $M_i(\vec{x})$ and $M_j(\vec{x})$ represent the responses of Mexican-Hat wavelets at the image-location $\vec{x}$ for scales *i* and *j*, respectively. For the image *A*, the wavelet response $M_i(\vec{x})$ is given by:

$$M_i(\vec{x}) = \langle (2^{-i}\psi(2^{-i} \cdot \vec{x})); A \rangle \qquad (12)$$

  where <;> denotes the convolution of its operands. The normalizing constant γ is given by $\gamma = 2^{-(i-j)}$, the operator $|\cdot|$ returns the absolute value of its parameter, and the $\psi(\vec{x})$ represents the response of the Mexican-Hat mother wavelet, and is defined as:

$$\psi(\vec{x}) = (2 - |\vec{x}|^2) \exp(-\frac{\vec{x}^2}{2}) \qquad (13)$$

- Determine points of local maximum of $P_{ij}(\cdot)$. These points correspond to the set of potential feature points.

- Accept a point of local maximum in $P_{ij}(\cdot)$ as a feature-point if the variance of the image-pixels in the neighborhood of the point is higher than a threshold. This criterion eliminates suspicious local maximum in featureless regions of the image.

The column-positions and row-positions of the resulting feature points are concatenated to form a string of digits, and then encrypted to generate the image signature. It is not hard to imagine that the file constructed in this way can have a smaller size compared to that constructed by recording the block histogram.

In order to determine whether an image *A* is authentic with another known image *B*, the feature set S_A of *A* is computed. The feature set S_A is then compared with the feature set S_B of *B* that is decrypted from the signature of *B*. The following rules are adopted to authenticate the image *A*.

- Verify that each feature location is present both in S_B and in S_A.
- Verify that no feature location is present in S_A but absent in S_B.
- Two feature points with coordinates and are said to match if:

$$|\vec{x} - \vec{y}| < 2 \qquad (14)$$

Edge-Based Methods

The edges in an image are the boundaries or contours where the significant changes occur in some physical aspects of an image, such as the surface reflectance, illumination, or the distances of the visible surfaces from the viewer. Edges are kinds of strong content features for an image. However, for common picture formats, coding edges value and position produces a huge overhead. One way to resolve this problem is to use a binary map to represent the edge. For example, Li, Lou and Liu (2003) used a binary map to encode the edges of an image in their watermarking-based image authentication scheme. It should be concerned that edges (both their position and value, and also the resulting binary image) might be modified if high compression ratios are used. Consequently, the success of using edges as the authentication code is greatly dependent on the capacity of the authentication system to discriminate the differences the edges produced by content-preserving manipulations from those content-changing manipulations. Queluz (2001) proposed an algorithm for edges extraction and edges integrity evaluation.

The block diagram of the edge extraction process of Queluz's method is shown as Figure 8. The gradient is first computed at each pixel position with an edge extraction operator. The result is then compared with an image-dependent threshold obtained from the image gradient histogram to obtain a binary image marking edge and no-edge pixels. Depending on the specifications for label size, the bit-map could be sub-sampled with the purpose of reducing its spatial resolution. Finally, the edges of the bit-map are encoded (compressed).

Edges integrity evaluation process is shown as Figure 9. In the *edges difference computation* block, the suspicious error pixels that have differences between the original and computed edge bit-maps and a certitude value associated with each error pixel are produced. These suspicious error pixels are evaluated in an *error relaxation* block. This is done by iteratively changing low certitude errors to high certitude errors if necessary, until no further change occurs. At the end, all high certitude errors are considered to be true errors and low certitude errors are eliminated. After error relaxation, the maximum connected region is computed according to a predefined threshold.

A similar idea was also proposed by Dittmann, Steinmetz and Steinmetz (1999). The feature-extraction process starts with extracting the edge characteristics C_I of the image *I* with the Canny edge detector *E* (Canny, 1986). The

Figure 8. Process of edge extraction proposed by Queluz (2001)

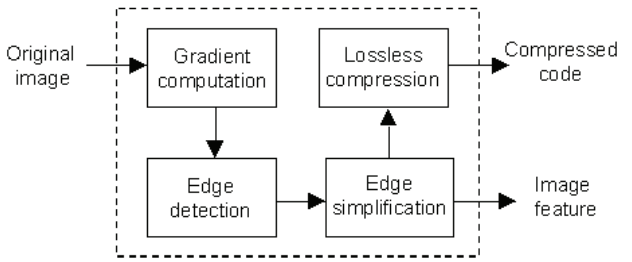
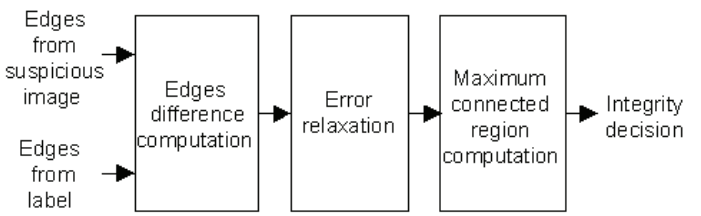


Figure 9. Process of edges integrity evaluation proposed by Queluz (2001)



C_I is then transformed to a binary edge pattern EP_{CI}. The variable length coding is then used to compress EP_{CI} into a feature code. This process is formulated as follows:

- Feature extraction: $C_I = E(I)$;
- Binary edge pattern: $EP_{CI} = f(C_I)$;
- Feature code: $VLC(EP_{CI})$.

The verification process begins with calculating the actual image edge characteristic C_T and the binary edge pattern EP_{CT}. The original binary edge pattern EP_{CI} is obtained by decompressing the received $VLC(EP_{CI})$. The EP_{CI} and CP_{CT} are then compared to obtain the error map. These steps can also be formulated as follows:

- Extract feature: $C_T = E(T)$, $EP_{CT} = f(C_T)$;
- Extract the original binary pattern: EP_{CI} = Decompress($VLC(EP_{CI})$);
- Check $EP_{CI} = EP_{CT}$.

Mean-Based Methods

Using local mean as the image feature may be the simplest and most practical way to represent the content character of an image. For example, Lou and Liu (2000) proposed an algorithm to generate a mean-based feature code. Figure 10 shows the process of feature code generation. The original image is first divided into non-overlapping blocks. The mean of each block is then calculated and quantized according to a predefined parameter. All the calculated results are then encoded (compressed) to form the authentication code. Figure 11 shows an example of this process. Figure 11(a) is a 256×256 gray image, and is used as the original image. It is first divided into 8×8 non-overlapping blocks. The mean of each block is then computed and is shown as Figure 11(b). Figure 11(c) also

shows the 16-step quantized block-means of Figure 11(b). The quantized block-means are further encoded to form the authentication code. It should be noted that Figure 11(c) is visually close to Figure 11(b). It means that the feature of the image is still preserved even though only the quantized block-means are encoded.

The verification process starts with calculating the quantized block-means of the received image. The quantized code is then compared with the original quantized code by using a sophisticated comparison algorithm. A binary error map is then produced as an output, with "1" denoting *match* and "0" denoting *mismatch*. The verifier can thus tell the possibly tampered blocks by inspecting the error map. It is worth mentioning that the quantized block-means can be used to repair the tampered blocks. This feasibility is attractive in the applications of the real-time image such as the video.

A similar idea was adopted in the process of generating the AIMAC (Approximate Image Message Authentication Codes) (Xie, Arce & Graveman, 2001). In order to construct a robust IMAC, an image is divided into non-overlapping 8×8 blocks, and the block mean of each block is computed. Then the most significant bit (MSB) of each block mean is extracted to form a binary map. The AIMAC is then generated according to this binary map. It should be noted that the histogram of the pixels in each block should be adjusted to preserve a gap of 127 gray levels for each block mean. In such a way, the MSB is robust enough to distinguish content-preserving manipulations from content-changing manipulations. This part has a similar effectiveness to the sophisticated comparison part of the algorithm proposed by Lou and Liu (2000).

Relation-Based Methods

Unlike the methods introduced above, relation-based methods divide the original image into non-overlapping blocks, and use the relation between blocks as the feature code. The method proposed by Lin and Chang (1998, 2001) is called SARI. The feature code in SARI is generated to survive the JPEG compression. To serve this purpose, the process of the feature code generation starts with dividing the original image into 8×8 non-overlapping blocks. Each block is then DCT transformed. The transformed DCT blocks are further grouped into two non-overlapping sets. There are equal numbers of DCT blocks in each set (i.e., there are $N/2$ DCT blocks in each set if the original image is divided into N blocks). A secret key-dependent mapping function then one-to-one maps each DCT block in one set into another DCT block in the other set, and generates $N/2$ DCT block pairs. For each block pair, a number of DCT coefficients are then selected and compared. The feature code is then generated by comparing the corresponding coefficients of the paired blocks. For example, if the coefficient in the first DCT block is greater than the coefficient in the second DCT block, then code is generated as "1". Otherwise, a "0" is generated. The process of generating the feature code is illustrated as Figure 12.

To extract the feature code of the received image, the same secret key should be applied in the verification process. The extracted feature code

Figure 10. Process of generation of image feature proposed by Lou and Liu (2000)

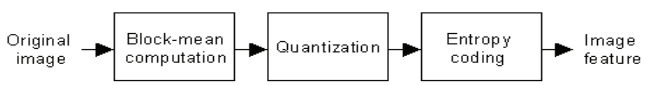
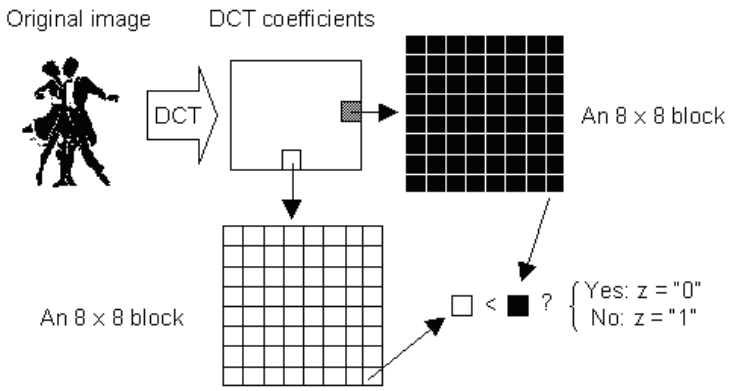


Figure 11. (a) Original image, (b) Map of block-means, (c) Map of 16-step quantized block-means

Figure 12. Feature code generated with SARI authentication scheme



is then compared with the original feature code. If either block in each block pair has not been maliciously manipulated, the relation between the selected coefficients is maintained. Otherwise, the relation between the selected coefficients may be changed.

It can be proven that the relationship between the selected DCT coefficients of two given image blocks is maintained even after the JPEG compression by using the same quantization matrix for the whole image. Consequently, SARI authentication system can distinguish JPEG compression from other malicious manipulations. Moreover, SARI can locate the tampered blocks because it is a block-wise method.

Structure-Based Methods

Lu and Liao (2000, 2003) proposed another kind of method to generate the feature code. The feature code is generated according to the structure of the image content. More specifically, the content structure of an image is composed of parent-child pairs in the wavelet domain. Let $w_{s,o}(x, y)$ be a wavelet coefficient at the scale s. Orientation o denotes horizontal, vertical, or diagonal direction. The inter-scale relationship of wavelet coefficients is defined for the parent node $w_{s+1,o}(x, y)$ and its four children nodes $w_{s,o}(2x+i, 2y+j)$ as either $|w_{s+1,o}(x, y)| \geq |w_{s,o}(2x+i, 2y+j)|$ or $|w_{s+1,o}(x, y)| \leq |w_{s,o}(2x+i, 2y+j)|$, where $0 \leq i, j \leq 1$.

The authentication code is generated by recording the parent-child pair that satisfies $\|w_{s+1,o}(x, y)| - |w_{s,o}(2x+i, 2y+j)\| > \rho$, where $\rho > 0$. Clearly, the threshold ρ is used to determine the size of the authentication code, and plays a trade-off role between robustness and fragility. It is proven that the inter-scale relationship is difficult to be destroyed by content-preserving manipulations and is hard to be preserved by content-changing manipulations.

DESIGN ISSUES

Digital signature-based image authentication is an important element in the applications of image communication. Usually, the content verifiers are not the creator or the sender of the original image. That means the original image is not available during the authentication process. Therefore, one of the fundamental requirements for digital signature-based image authentication schemes is *blind authentication*, or *obliviousness*, as it is sometimes called. Other requirements depend on the applications that may be based on strict authentication or non-strict authentication. In this section, we will discuss some issues about designing effective digital signature-based image authentication schemes.

Error Detection

In some applications, it is proper if modification of an image can be detected by authentication schemes. However, it is beneficial if the authentication schemes are able to detect or estimate the errors so that the distortion can be compensated or even corrected. Techniques for error detection can be categorized into two classes according to the applications of image authentication; namely, error type and error location.

Error Type

Generally, strict authentication schemes can only determine whether the content of the original image is modified. This also means that they are not able to differentiate the types of distortion (e.g., compression or filtering). By contrast, non-strict authentication schemes tend to tolerate some form of errors. The key to developing a non-strict authentication scheme is to examine what the digital signature should protect. Ideally, the authentication code should protect the message conveyed by the content of the image, but not the particular representation of that content of the image. Therefore, the authentication code can be used to verify the authenticity of an image that has been incidentally modified, leaving the value and meaning of its contents unaffected. Ideally, one can define an authenticity versus modification curve such as the method proposed by Schneider and Chang (1996) to achieve the desired authenticity. Based on the authenticity versus modification curve, authentication is no longer a yes-or-no question. Instead, it is a continuous interpretation. An image that is bit by bit identical to the original image has an authenticity measure of 1.0 and is considered to be completely authentic. An image that has nothing in common with the original image has an authenticity measure of 0.0 and is considered unauthentic. Each of the other images would have authenticity measure between the range (0.0, 1.0) and be partially authentic.

Error Location

Another desirable requirement for error detection in most applications is errors localization. This can be achieved by block-oriented approaches. Before transmission, an image is usually partitioned into blocks. The authentication code of each block is calculated (either for strict or non-strict authentication). The authentication codes of the original image are concatenated, signed, and transmitted as a separate file. To locate the

distorted regions during the authenticating process, the received image is partitioned into blocks first. The authentication code of each block is calculated and compared with the authentication code recovered from the received digital signature. Therefore, the smaller the block size is, the better the localization accuracy is. However, the higher accuracy is gained at the expense of the larger authentication code file and the longer process of signing and decoding. The trade-off needs to be taken into account at the designing stage of an authentication scheme.

Error Correction

The purpose of error correction is to recover the original images from their manipulated version. This requirement is essential in the applications of military intelligence and motion pictures (Dittmann, Steinmetz & Steinmetz, 1999; Queluz, 2001). Error correction can be achieved by means of error correction code (ECC) (Lin & Costello, 1983). However, encrypting ECC along with feature code may result in a lengthy signature. Therefore, it is more advantageous to enable the authentication code itself to be the power of error correction. Unfortunately, the authentication code generated by strict authentication schemes is meaningless and cannot be used to correct the errors. Compared to strict authentication, the authentication code generated by non-strict authentication schemes is potentially capable of error correction. This is because the authentication code generated by the non-strict authentication is usually derived from the image feature and is highly content dependent.

An example of using authentication code for image error correction can be found in Xie, Arce and Graveman (2001). This work uses quantized image gray values as authentication code. The authenticated code is potentially capable of error correcting since image features are usually closely related to image gray values. It should be noted that the smaller the quantization step is, the better the performance of error correction is. However, a smaller quantization step also means a longer signature. Therefore, trade-off between the performance of error correction and the length of signature has to be made as well. This is, without doubt, an acute challenge, and worth further researching.

Security

With the protection of public-key encryption, the security of the digital signature-based image authentication is reduced to the security of the image digest function that is used to produce the authentication code. For strict authentication, the attacks on hash functions can be grouped into two categories: brute-force attacks and cryptanalysis attacks.

Brute-Force Attacks

It is believed that, for a general-purpose secure hash code, the strength of a hash function against brute-force attacks depends solely on the length of the hash code produced by the algorithm. For a code of length n, the level of effort required is proportional to $2^{n/2}$. This is also known as *birthday attack*. For example, the length of the hash code of MD5 (Rivest, 1992) is 128 bits. If an attacker has 2^{64} different samples, he or she has more than 50% of chances to find the same hash code. In other words, to create a fake image that has the same hash result as the original image, an attacker only needs to prepare 2^{64} visually equivalent fake images. This can be accomplished by first creating a fake image and then varying the least significant bit of each of 64 arbitrarily chosen pixels of the fake image. It has been proved that we could find a collision in 24 days by using a $10 million collision search machine for MD5 (Stallings, 2002). A simple solution to this problem is to use a hash function to produce a longer hash code. For example, SHA-1 (NIST FIPS PUB 180, 1993) and RIPEMD-160 (Stall-

ings, 2002) can provide 160-bit hash code. It is believed that over 4,000 years would be required if we used the same search machine to find a collision (Oorschot & Wiener, 1994). Another way to resolve this problem is to link the authentication code with the image feature such as the strategy adopted by non-strict authentication.

Non-strict authentication employs image feature as the image digest. This makes it harder to create enough visually equivalent fake images to forge a legal one. It should be noted that, mathematically, the relationship between the original image and the authentication code is many-to-one mapping. To serve the purpose of error tolerance, non-strict authentication schemes may have one authentication code corresponding to more images. This phenomenon makes non-strict authentication approaches vulnerable and remains as a serious design issue.

Cryptanalysis Attacks

Cryptanalysis attacks on digest function seek to exploit some property of the algorithm to perform some attack rather than an exhaustive search. Cryptanalysis on the strict authentication scheme is to exploit the internal structure of the hash function. Therefore, we have to select a secure hash function that can resist cryptanalysis performed by attackers. Fortunately, so far, SHA-1 and RIPEMD-160 are still secure for various cryptanalyses and can be included in strict authentication schemes. Cryptanalysis on non-strict authentication has not been defined so far. It may refer to the analysis of key-dependent digital signature-based schemes. In this case, an attacker tries to derive the secret key from multiple feature codes, which is performed in a SARI image authentication system (Radhakrisnan & Memon, 2001). As defined in the second section, there is no secret key involved in a digital signature-based authentication scheme. This means that the secrecy of the digital signature-based authentication schemes depends on the robustness of the algorithm itself and needs to be noted for designing a secure authentication scheme.

CONCLUSION

With the advantages of the digital signature (Agnew, Mullin & Vanstone, 1990; ElGamal, 1985; Harn, 1994; ISO/IEC 9796, 1991; NIST FIPS PUB, 1993; Nyberg & Rueppel, 1994; Yen & Laih, 1995), digital signature-based schemes are more applicable than any other schemes in image authentication. Depending on applications, digital signature-based authentication schemes are divided into strict and non-strict categories and are described in great detail in this chapter. For strict authentication, the authentication code derived from the calculation of traditional hash function is sufficiently short. This property enables fast creation of the digital signature. In another aspect, the arithmetic-calculated hash is very sensitive to the modification of image content. Some tiny changes to a single bit in an image may result in a different hash. This results in that strict authentication can provide binary authentication (i.e., yes or no). The trustworthy camera is a typical example of this type of authentication scheme.

For some image authentication applications, the authentication code should be sensitive for content-changing modification and can tolerate some content-preserving modification. In this case, the authentication code is asked to satisfy some basic requirements. Those requirements include locating modification regions and tolerating some forms of image processing operations (e.g., JPEG lossy compression). Many non-strict authentication techniques are also described in this chapter. Most of them are designed to employ a special-purpose authentication code to satisfy those basic requirements shown above. However, few of them are capable of recovering some certain errors. This special-purpose authentication code may be the modern and useful aspect for non-strict authentication.

Under the quick evolution of image processing techniques, existing digital signature-based image authentication schemes will be further improved to meet new requirements. New requirements pose new challenges for designing effective digital signature-based authentication schemes. These challenges may include using large-size authentication code and tolerating more image-processing operations without compromising security. This means that new approaches have to balance the trade-off among these requirements. Moreover, more modern techniques combining the watermark and digital signature techniques may be proposed for new image authentication generations. Those new image authentication techniques may result in some changes of the watermark and digital signature framework, as demonstrated in Sun and Chang (2002), Sun, Chang, Maeno and Suto (2002a, 2002b) and Lou and Sung (to appear).

REFERENCES

Agnew, G.B., Mullin, R.C., & Vanstone, S.A. (1990). Improved digital signature scheme based on discrete exponentiation. *IEEE Electronics Letters, 26,* 1024-1025.

Bhattacharjee, S., & Kutter, M. (1998). Compression tolerant image authentication. *Proceedings of the International Conference on Image Processing, 1,* 435-439.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8*(6), 679-698.

Diffie, W., & Hellman, M.E. (1976). New directions in cryptography. *IEEE Transactions on Information Theory, IT-22*(6), 644-654.

Dittmann, J., Steinmetz, A., & Steinmetz, R. (1999). Content-based digital signature for motion pictures authentication and content-fragile watermarking. *Proceedings of the IEEE International Conference On Multimedia Computing and Systems, 2,* 209-213.

ElGamal, T. (1985). A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory, IT-31*(4), 469-472.

Friedman, G.L. (1993). The trustworthy digital camera: Restoring credibility to the photographic image. *IEEE Transactions on Consumer Electronics, 39*(4), 905-910.

Harn, L. (1994). New digital signature scheme based on discrete logarithm. *IEE Electronics Letters, 30*(5), 396-398.

ISO/IEC 9796. (1991). Information technology security techniques digital signature scheme giving message recovery. *International Organization for Standardization.*

Li, C.-T., Lou, D.-C., & Chen, T.-H. (2000). Image authentication via content-based watermarks and a public key cryptosystem. *Proceedings of the IEEE International Conference on Image Processing, 3,* 694-697.

Li, C.-T., Lou, D.-C., & Liu, J.-L. (2003). Image integrity and authenticity verification via content-based watermarks and a public key cryptosystem. *Journal of the Chinese Institute of Electrical Engineering, 10*(1), 99-106.

Lin, C.-Y., & Chang, S.-F. (1998). A robust image authentication method surviving JPEG lossy compression. *SPIE storage and retrieval of image/video databases.* San Jose.

Lin, C.-Y., & Chang, S.-F. (2001). A robust image authentication method distinguishing JPEG Compression from malicious manipulation. *IEEE Transactions on Circuits and Systems of Video Technology, 11*(2), 153-168.

Lin, S., & Costello, D.J. (1983). *Error control coding: Fundamentals and applications.* NJ: Prentice-Hall.

Lou, D.-C., & Liu, J.-L. (2000). Fault resilient and compression tolerant digital signature for image authentication. *IEEE Transactions on Consumer Electronics, 46*(1), 31-39.

Lou, D.-C., & Sung, C.-H. (to appear). A steganographic scheme for secure communications based on the chaos and Euler theorem. *IEEE Transactions on Multimedia.*

Lu, C.-S., & Liao, M.H.-Y. (2000). Structural digital signature for image authentication: An incidental distortion resistant scheme. *Proceedings of Multimedia and Security Workshop at the ACM International Conference On Multimedia,* pp. 115-118.

Lu, C.-S., & Liao, M.H.-Y. (2003). Structural digital signature for image authentication: An incidental distortion resistant scheme. *IEEE Transactions on Multimedia, 5*(2), 161-173.

NIST FIPS PUB. (1993). *Digital signature standard.* National Institute of Standards and Technology, U.S. Department of Commerce, DRAFT.

NIST FIPS PUB 180. (1993). *Secure hash standard.* National Institute of Standards and Technology, U.S. Department of Commerce, DRAFT.

Nyberg, K., & Rueppel, R. (1994). Message recovery for signature schemes based on the discrete logarithm problem. *Proceedings of Eurocrypt'94,* 175-190.

Oorschot, P.V., & Wiener, M.J. (1994). Parallel collision search with application to hash functions and discrete logarithms. *Proceedings of the Second ACM Conference on Computer and Communication Security,* 210-218.

Queluz, M.P. (2001). Authentication of digital images and video: Generic models and a new contribution. *Signal Processing: Image Communication, 16,* 461-475.

Radhakrisnan, R., & Memon, N. (2001). On the security of the SARI image authentication system. *Proceedings of the IEEE International Conference on Image Processing, 3,* 971-974.

Rivest, R.L. (1992). The MD5 message digest algorithm. *Internet Request For Comments 1321.*

Rivest, R.L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM, 21*(2), 120-126.

Schneider, M., & Chang, S.-F. (1996). Robust content based digital signature for image authentication. *Proceedings of the IEEE International Conference on Image Processing, 3,* 227-230.

Stallings, W. (2002). *Cryptography and network security: Principles and practice* (3rd ed.). New Jersey: Prentice-Hall.

Sun, Q., & Chang, S.-F. (2002). Semi-fragile image authentication using generic wavelet domain features and ECC. *Proceedings of the 2002 International Conference on Image Processing, 2,* 901-904.

Sun, Q., Chang, S.-F., Maeno, K., & Suto, M. (2002a). A new semi-fragile image authentication framework combining ECC and PKI infrastructures. *Proceedings of the 2002 IEEE International Symposium on Circuits and Systems, 2,* 440-443.

Sun, Q., Chang, S.-F., Maeno, K., & Suto, M. (2002b). A quantitive semi-fragile JPEG2000 image authentication system. *Proceedings of the 2002 International Conference on Image Processing, 2,* 921-924.

Wallace, G.K. (1991, April). The JPEG still picture compression standard. *Communications of the ACM, 33,* 30-44.

Walton, S. (1995). Image authentication for a slippery new age. *Dr. Dobb's Journal, 20*(4), 18-26.

Xie, L., Arce, G.R., & Graveman, R.F. (2001). Approximate image message authentication

codes. *IEEE Transactions on Multimedia, 3*(2), 242-252.

Yen, S.-M., & Laih, C.-S. (1995). Improved digital signature algorithm. *IEEE Transactions on Computers, 44*(5), 729-730.

This work was previously published in Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property, edited by C.-S. Lu, pp. 207-230, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.11
A Stochastic Content-Based Image Retrieval Mechanism

Mei-Ling Shyu
University of Miami, USA

Shu-Ching Chen
Florida International University, USA

Chengcui Zhang
Florida International University, USA

ABSTRACT

Multimedia information, typically image information, is growing rapidly across the Internet and elsewhere. To keep pace with the increasing volumes of image information, new techniques need to be investigated to retrieve images intelligently and efficiently. Content-based image retrieval (CBIR) is always a challenging task. In this chapter, a stochastic mechanism, called Markov Model Mediator (MMM), is used to facilitate the searching and retrieval process for content-based image retrieval, which serves as the retrieval engine of the CBIR systems and uses stochastic-based similarity measures. Different from the common methods, our stochastic mechanism carries out the searching and similarity computing process dynamically, taking into consideration not only the image content features but also other characteristics of images such as their access frequencies and access patterns. Our experimental results demonstrate that the MMM mechanism together with the stochastic process can assist in retrieving more accurate results for user queries.

INTRODUCTION

The availability of today's digital devices and techniques offers people more opportunities than ever to create their own digital images. Moreover, the Internet has become the biggest platform to get, distribute and exchange digital image data. The rapid increase in the amount of image data and the inefficiency of traditional text-based image retrieval have created great demands for new approaches in image retrieval. As a consequence

of such fast growth of digital image databases, the development of efficient search mechanisms has become more and more important. Content-Based Image Retrieval (CBIR) emerges and is dedicated to tackling such difficulties. CBIR is an active research area where the image retrieval queries are based on the content of multimedia data.

In contrast to the text-based approach, CBIR operates on a totally different principle, i.e., to retrieve the stored images from a collection of images by comparing the features that were automatically extracted from the images themselves. Content-based image retrieval involves a matching process between a query image and the images stored in the database. The first step of the process involves extracting a feature vector for the unique characteristics of each image. The features used for retrieval can be either primitive or semantic, but the extraction process must be automatic. A quantified similarity value between two images is obtained by comparing their feature vectors. The commonly used image features include color, shape and texture. Queries are issued through query by image example (QBE), which can either be provided or constructed by users, or randomly selected from the image database. A lot of research work has been done, which resulted in a number of systems and techniques, both in the academic and commercial domains. For example, IBM's QBIC system (Faloutsos, 1994; Flickner, 1995) and Virage's VIR engine (*http://www.virage.com*) are two of the most notable commercial image retrieval systems, while VisualSEEk (Smith, 1996) and PhotoBook (Pentland, 1994) are well-known academic image retrieval systems.

For large image collections, traditional retrieval methods such as sequential searching do not work well since it is time expensive and tends to ignore the relationships among all images by only considering the relationship between the query image and a single image in the database. Various kinds of data structures, approaches and techniques have been proposed to manage image databases and hasten the retrieval process.

The first aim of this chapter is to take an overview of the currently available content-based image retrieval (CBIR) systems. Then, with the focus on the searching process, we present a conceptually coherent framework that adopts a stochastic mechanism called Markov Model Mediator (MMM) for the content-based image retrieval problem. With an explicit model of image query patterns, given the target image, the proposed framework carries out the searching and similarity computing process dynamically by taking into consideration not only the image content features but also their access frequencies and access patterns.

We begin with the literature review as well as the motivations of the proposed mechanism. Then, the key components of the MMM mechanism and the stochastic process for information retrieval are introduced. After that, our experiments in applying the MMM mechanism to content-based image retrieval are presented and the experimental results are also provided. The future trends are discussed as well. Finally, a brief conclusion is given.

BACKGROUND

The objective of a CBIR system is to offer the user an efficient way in finding and retrieving those images that are qualified for the matching criteria of the users' queries from the database. Most of the existing CBIR systems retrieve images in the following manner. First, they build the indexes based on the low-level features such as color, texture and shape for the images in the database. The corresponding indexes of a query image are also generated at the time the query is issued. Second, they search through the whole database and measure the similarity of each image to the query image. Finally, the results are returned in a sorted order of the similarity matching level.

Lots of approaches for retrieving images on the basis of color similarity have been described

in the literature, but most of them are actually variations of the same basic idea. The most commonly used matching technique is histogram intersection (Li, 2000). Variants of this technique are now used in a big proportion of the current CBIR systems. Methods of improving the original technique include the use of cumulative color histograms, combining histogram intersection with some element of spatial matching (Stricker, 1996) and the use of region-based color querying (Carson, 1997). As for texture similarity, the useful measures include the degree of *contrast*, *coarseness*, *directionality*, *regularity*, *periodicity*, *randomness*, etc. Alternative methods of texture analysis for retrieval include the use of Gabor filters (Ma, 1998) and fractals (Kaplan, 1998). Unlike texture, shape is a fairly well defined concept. Two main types of shape features are commonly used, namely (1) the *global* features such as aspect ratio, circularity and moment invariants and (2) the *local* features such as sets of consecutive boundary segments. Alternative methods proposed for shape matching include elastic deformation of templates (Zhong, 2000).

An impediment to research on CBIR is the lack of mapping between the high-level concepts and the low-level features. In order to overcome this problem and to better capture the subjectivity of human perception of the visual content, the concept of relevance feedback (RF) associated with CBIR was proposed in Rui (1997). Relevance feedback is an interactive process in which the user judges the quality of the retrieval results performed by the system by marking those images that the user perceives as truly relevant among the images retrieved by the system. This information is then used to refine the original query. However, even if the user provides a good initial query, it is still a problem of how to navigate through the image database.

No matter what information and what techniques are used for the construction of the image indexes, and no matter what similarity measurement strategies are employed, as far as the searching process is concerned, simple approaches such as sequential searching, are commonly put into operations to find the group of similar images for the queries. Such kinds of approaches may be adequate for small databases. However, as the scales and volumes of the databases increase considerably, they are deficient. Moreover, these approaches focus on only the relationship between the query image and the target image, neglecting the relationships among all the images within the database, which may result in inflexible and incomplete searching results.

There have been quite a few techniques being proposed and employed to alleviate the time consumption problem and to speed up the retrieval process, such as efficient indexing structures, compact representations, and pre-filtering techniques (Hafner, 1995). The QBIC system (Faloutsos, 1994; Flickner, 1995), for instance, uses the pre-filtering technique and the efficient indexing structure like R-trees to accelerate its searching performance. However, little has been done in considering the complicated relationships of the image objects to each other.

In this chapter, we present a content-based retrieval system that employs the Markov model mediator (MMM) mechanism to retrieve images, which functions as both the searching engine and image similarity arbitrator. In our previous studies, the MMM mechanism has been applied to multimedia database management (Shyu, 2000a, 2000c) and document management on the World Wide Web (WWW) (Shyu, 2000b, 2001a, 2001b). The MMM mechanism adopts the Markov model framework and the concept of the mediators. The Markov model is one of the most powerful tools available to scientists and engineers for analyzing complicated systems, whereas a mediator is defined as a program that collects and combines information from one or more sources, and finally yields the resulting information (Wiederhold, 1992). A Markov model consists of a number of states connected by transitions. The Markov property states that given the current state of the

system, the future evolution is independent of its history. In other words, the states represent the alternatives of the stochastic process and the transitions contain probabilistic and other data used to determine which state should be selected next. All the transitions $S_i \rightarrow S_j$ such that $Pr(S_j \mid S_i) > 0$ are said to be allowed, the rest are prohibited. Markov models have been used in many applications. Some well-known examples are Markov Random Field Models in Frank (1986), and Hidden Markov Models (HMMs) (Rabiner, 1986). Some research works have been done to integrate the Markov model into the field of image retrieval. Lin et al. (Lin, 1997) used a Markov model to combine the spatial and color information. In their approach, each image in the database is represented by a pseudo two-dimensional hidden Markov model (HMM) in order to adequately capture both the spatial and chromatic information about that image. Wolf (1997) used the hidden Markov model (HMM) to parse video data. In Naphade (2001), the hidden Markov model was employed to model the time series of the feature vector for the cases of events and objects in their probabilistic framework for semantic level indexing and retrieval.

Our proposed CBIR system employs the MMM mechanism as well as the stochastic-based similarity measures for dynamic content-based image retrieval, which retrieves images with respect to the user queries. Our method also builds an index vector for each image within the database, but unlike the common methods mentioned above, our method considers not only the relationship between the query image and the target image, but also the relationships among all images in the database. A stochastic process that takes into account the image content features and other characteristics (such as the access frequencies and access patterns) of the images is proposed. Several experiments are conducted and the experimental results demonstrate that the MMM mechanism together with the stochastic process can assist in retrieving more accurate results for user queries.

THE STOCHASTIC MODEL

Markov Model Mediator (MMM) Mechanism

Markov model mediator (MMM) is a probabilistic-based mechanism that adopts the Markov model framework and the mediator concept (Shyu, 2000a, 2000b, 2000c, 2001a, 2001b). In our CBIR system, each image database is modeled by an MMM. The MMM mechanism is defined as follows.

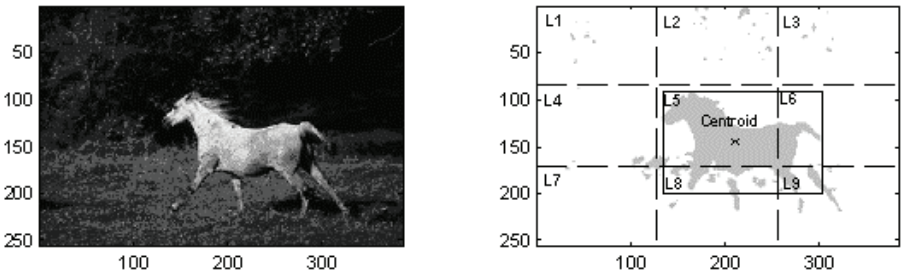
Definition 1: An MMM is represented by a 5-tuple $\lambda = (\mathcal{S}, \mathcal{F}, \mathcal{A}, \mathcal{B}, \Pi)$*, where* $\mathcal{S}$ *is a set of images called states;* $\mathcal{F}$ *is a set of distinct features of the images;* $\mathcal{A}$ *denotes the state transition probability distribution, where each entry (i, j) actually indicates the relationship between images i and j;* $\mathcal{B}$ *is the observation symbol probability distribution; and* Π *is the initial state probability distribution.*

Each image database is modeled by an MMM, where $\mathcal{S}$ consists of all the images in the image database and $\mathcal{F}$ includes all the distinct features for the images in $\mathcal{S}$. The elements in $\mathcal{S}$ and $\mathcal{F}$ determine the dimensions of $\mathcal{A}$ and $\mathcal{B}$. If there are totally n images in $\mathcal{S}$, and the number of distinct features in $\mathcal{F}$ is m, then the dimensions of $\mathcal{A}$ is $n \times n$, while $\mathcal{B}$ has the size of $n \times m$. The relationships of the images are modeled by the sequences of the MMM states connected by transitions. A training data set consisting of the access patterns and access frequencies of the queries issued to the database is used to train the model parameters for an MMM.

Formulation of the Model Parameters

Each MMM has three probability distribution matrixes: $\mathcal{A}$, $\mathcal{B}$, and Π. These matrixes are criti-

Figure 1. Object locations and their corresponding regions



cal for the stochastic process and can be obtained from the training data set.

Matrix B: The Observation Symbol Probability Distribution

The observation symbol probability $\mathcal{B}$ denotes the probability of observing an output symbol from a state. Here, the observed output symbols represent the distinct features of the images and the states represent the images in the database. Since an image has one or more features and one feature can appear in multiple images, the observation symbol probabilities show the probabilities that a feature is observed from a set of images.

In this study, we consider the following features: color information and object location information for the images in the image database. Since the color feature is closely associated with image scenes and it is more robust to changes due to scaling, orientation, perspective and occlusion of images, it is the most widely used visual feature in image retrieval (Ma, 1999). Humans perceive a color as a combination of three stimuli, R (red), G (Green), and B (Blue), which form a color space. Separating chromatic information and luminance information can generate more color spaces such as RGB, YIQ, YUV, CIE LAB, CIE LUV, and HSV. None of them can be used for all applications (Androutsos, 1999; Aslandogan, 1999; Cheng, 2001, 2000; Ma, 1999; Rui, 1999). RGB is the most commonly used color space primarily because color image acquisition and recording hardware are designed for this space. However, the problem of this space is the close correlation among the three components, which means that all three components will change as the intensity changes. This is not good for color analysis. YIQ and YUV are used to represent the color information in TV signals in color television broadcasting. CIE LAB and CIE LUV are often used in measuring the distance between two colors because of their perceptual uniformity. However, its transformation from the RGB space is computationally intensive and dependent on a reference white.

In our CBIR system, color information is obtained for each image from its HSV color space. The whole color space is divided into 12 subspaces according to the combinations of different ranges of hue (H), saturation (S), and intensity values (V). The HSV color space is chosen for two reasons. First, it is perceptual, which makes HSV a proven color space particularly amenable to color image analysis (Androutsos, 1999; Cheng, 2001, 2000). Secondly, the benchmark results in Ma (1999) shows that the color histogram in the HSV color space performs the best. For information of object locations, the SPCPE algorithm proposed in Sista (1999) and Chen (2000) is used.

Figure 2. Three sample images (Img1 – Img3)

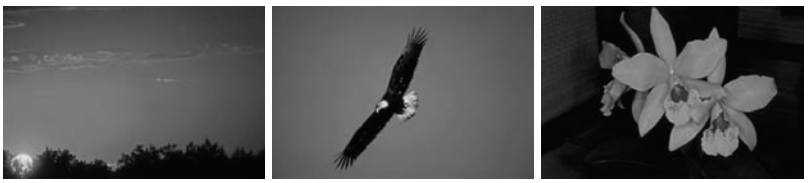


Table 1. $\mathcal{BB}$ matrix: Image feature vectors of sample images

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img1	0.11	0	0.83	0.06	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
img2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
img3	0.53	0	0	0.30	0	0	0.17	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table 2. $\mathcal{B}$ matrix: Normalized image feature vectors of sample images

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img1	0.06	0	0.41	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0.25	0
img2	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0
img3	0.26	0	0	0.15	0	0	0.09	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0

The minimal bounding rectangle (MBR) concept in R-tree (Guttman, 1984) is adopted so that each object is covered by a rectangle. The centroid point of each object is used for space reasoning so that any object is mapped to a point object. When these features are integrated into the queries, the semantic level meaning in the users' queries can be captured.

In our experiments, each image has a feature vector of 21 elements. Within the 21 features, 12 are for color descriptions and nine are for location descriptions. The color features considered are 'black,' 'white' (w), 'red,' 'red-yellow' (ry), 'yellow' (y), 'yellow-green' (yg), 'green,' 'green-blue' (gb), 'blue' (b), 'blue-purple' (bp), 'purple' (p), and 'purple-red' (pr). Colors with a number of pixels less than 5% of the total number of pixels are regarded as non-important and the corresponding positions in the feature vector have a value of 0. Otherwise, we put the corresponding percentage of that color component to that position. As for the location descriptions, each image is divided into 3×3 equal-sized regions. The image can be divided into a coarser or finer set of regions if necessary. As shown in Figure 1, the nine regions are ordered from left to right and top to bottom: L1, L2, L3, L4, L5, L6, L7, L8, and L9. When there is an object in the image whose centroid falls into one of the nine regions, the value 1 is assigned to that region. Objects with their areas less than 8% of the total area are ignored.

In order to capture the appearance of a feature in an image, we define a temporary matrix ($\mathcal{BB}$) whose rows are all the distinct images and columns are all the distinct features, where the value in the $(p, q)^{th}$ entry is greater than zero if feature

q appears in image p, and 0 otherwise. Then, the observation symbol probability distribution $\mathcal{B}$ can be obtained via normalizing $\mathcal{BB}$ per row. In other words, the sum of the probabilities that the features are observed from a given image should be 1.

Matrix $\mathcal{BB}$ consists of image feature vectors for all images. Figure 2 gives three example images and Table 1 illustrates their associated feature vectors. The observation symbol probability distribution $\mathcal{B}$ can be obtained via normalizing $\mathcal{BB}$ per row as shown in Table 2. In other words, the sum of the probabilities that the features are observed from a given image should be 1. We consider that the color and location information are of equal importance, such that the sum of observed probability of color features should be equal to that of location features (0.5 each).

Training Data Set

A set of training data is used to generate the training traces that are the central part of the stochastic process. Definition 2 gives the information that is available in the training data set. Based on the information in the training data set, we calculate the relative affinity measurements of the images in the image database (as shown in Definition 3).

Definition 2: The training data set consists of the following information:

• *The value N that indicates the number of images in the image database d.*

• *A set of queries* $Q = \{q_1, q_2, \ldots, q_q\}$ *that are issued to the database in a period of time. Each query shows the access patterns and access frequencies of the images. Let* $use_{m,k}$ *denote the usage pattern of image m with respect to query* q_k *per time period, where the value of* $use_{m,k}$ *is 1 when m is accessed by* q_k*, and 0 otherwise. The value of* $access_k$ *denotes the access frequency of query* q_k *per time period.*

Definition 3: The relative affinity measurements indicate how frequently two images are accessed together. The relative affinity relationship between two images m and n is defined as follows:

$$aff_{m,n} = \sum_{k=1}^{q} use_{m,k} \times use_{n,k} \times access_k \qquad (1)$$

Table 3 gives eight example queries issued to the image database with their corresponding access frequencies. The access patterns of the three sample images in the database versus the eight example queries are shown in Table 4. In this table, the entry $(i, j) = 1$ indicates that the i^{th} image is accessed by the j^{th} query (i.e., q_j).

Matrix A: The State Transition Probability Distribution

The state transition probability distribution (matrix A) is constructed by having $a_{m,n}$ be the element in the $(m, n)^{th}$ entry in A, where:

$$a_{m,n} = \frac{f_{m,n}}{f_m} \qquad (2)$$

$$f_{m,n} = \frac{aff_{m,n}}{\sum_{m \in d} \sum_{n \in d} aff_{m,n}} \qquad (3)$$

$$f_m = \sum_n f_{m,n} \qquad (4)$$

In this formulation, $f_{m,n}$ is the joint probability that refers to the fraction of the relative affinity of images m and n in the database d with respect to the total relative affinity for all the images in d, f_m is the marginal probability, and $a_{m,n}$ is the conditional probability that refers to the state transition probability for an MMM.

Matrix Π: *The Initial State Probability Distribution*

The preference of the initial states for queries can be obtained from the training traces. For any image

Table 3. Eight example queries and their frequencies ($access_k$)

Query Type	Feature Required	Frequency
q1	black / L1	1200
q2	blue	1500
q3	w hite / red / L 5	2500
q4	yellow / L5	1750
q5	green / gb	1250
q6	purple / L9	2220
q7	ry / L 5	1870
q8	bp	1345

Table 4. The access patterns of the sample images

	q1	q2	q3	q4	q5	q6	q7	q8
img1	1	0	1	1	0	0	1	0
img2	0	1	1	1	0	0	1	0
img3	1	0	1	1	1	0	1	0
... ...	... ...	... ...	... ...	... ...	... ...	... ...	... ...	... ...	... ...

$m \in d$, the initial state probability is defined as the fraction of the number of occurrences of image m with respect to the total number of occurrences for all the images in d from the training traces.

$$\Pi = \{\pi_m\} = \frac{\sum_{k=1}^{q} use_{m,k}}{\sum_{l=1}^{N}\sum_{k=1}^{q} use_{l,k}} \qquad (5)$$

Stochastic Process for Information Retrieval

The need for efficient information retrieval from databases is strong. Usually the cost for query processing is expensive and time-consuming. Meanwhile, the results may not be very satisfactory. Probabilistic models offer a way to perform the searching process more efficiently and accurately. We capture the most matched images through a dynamic programming algorithm that conducts a stochastic process in calculating the current edge weights and the cumulative edge weights.

Assume N is the total number of images in the databases and each query is denoted as *query* = $\{o_1, o_2, \ldots, o_T\}$, where T is the total number of features requested in the query. We define the edge weights and the cumulative edge weights as follows.

Definition 4: $W_t(i, j)$ is defined as the edge weight of the edge $S_i \rightarrow S_j$ at the evaluation of the t^{th} *feature* (o_t) *in the query, where* $1 \le i, j \le N$ *and* $1 \le t \le T$.

Definition 5: $D_t(i, j)$ is defined as the cumulative edge weight of the edge $S_i \rightarrow S_j$ at the evaluation of the t^{th} *feature* (o_t) *in the query, where* $1 \le i, j \le N$ *and* $1 \le t \le T$.

Based on Definitions 4 and 5, the *dynamic programming algorithm* is given as follows:

At $t = 1$, we define

$$W_1(i, j) = \begin{cases} \pi_{S_i} b_{S_i}(o_1) & i = j \\ 0 & otherwise \end{cases} \quad (6)$$

$$D_1(i, j) = W_1(i, j) \quad (7)$$

The values of $W_{t+1}(i, j)$ and $D_{t+1}(i, j)$, where $1 \leq t \leq T - 1$, are calculated using the values of $W_t(i, j)$ and $D_t(i, j)$ as follows:

$$W_{t+1}(i, j) = \max_k (D_t(k, i) a_{S_i, S_j}) b_{S_j}(o_{t+1}) \quad (8)$$

$$D_{t+1}(i, j) = (\max_k D_t(k, i)) + W_{t+1}(i, j) \quad (9)$$

As we mentioned before, $\mathcal{A} = \{a_{S_i, S_j}\}$ denotes the states transition probability distribution, $\mathcal{B} = \{b_{S_j}(o_k)\}$ is the observation symbol probability distribution, and $\Pi = \{\pi_{S_i}\}$ is the initial state probability distribution.

The image retrieval steps using the dynamic programming algorithm in the stochastic process are shown in Table 5. As can be seen from the result, our method can give a good ordering using this stochastic process.

In Step 3, since we already obtained matrices $W_1(i, j)$ and $D_1(i, j)$ from Step 2, and the second feature o_2 is known, the content of $W_2(i, j)$ and $D_2(i, j)$ can be determined. The value of $D_t(i, j)$ (obtained in Step 4) represents the cumulative edge weight for the joint event that $\{o_1, o_2, \ldots, o_T\}$ is observed. In the filtering step (Step 6), $D_T(q, j)$ together with $sumD_t(j)$ (where $1 \leq t \leq T$) are used as the filter to retrieve the candidate images with respect to query image q. In this step, when there are images with identical $D_T(q, j)$ values, we go to the matrix $sumD_T(j)$ to find different values to order them. If we fail to order them by the values in $sumD_T(j)$, we have to trace down to the next matrix $sumD_{T-1}(j)$ and continue the process until we reach the first matrix $sumD_1(j)$. We also take into consideration the characteristics of $sumW_t(j)$. From our observations, if the j^{th} image does not have the t^{th} feature in the query, the value of $sumW_t(j)$ would be zero. Taking advantage of this characteristic, we can exclude some of the images that do not have any feature desired in the query from the final result. Therefore, in Step 7, we use $W_T(q, j) + W_T(j, q)$ to reflect the possibility that the j^{th} image matches the issued query. In other words, it indicates the matching percentage of the j^{th} image in the image database to the query image q with respect to the features $\{o_1, o_2, \ldots, o_T\}$.

EXPERIMENTS

Experimental Image Database System

In our image database, there are 1,500 color images of various dimensions that are used to carry out the experiments. With the purpose of supporting semantic level meaning in the users' queries, both the color information and object location information are considered in our experiments. In addition, the query-by-example strategy is used for query issuing in our experiments. Based on the training data set of this image database, first we need to construct the model parameters for the MMM mechanism for the database.

Constructions of the Model Parameters

Each MMM has three probability distributions ($\mathcal{A}$, $\mathcal{B}$, and Π). The state transition probability distribution $\mathcal{A}$ can be obtained according to Equations (1) to (4). In order to calculate $\mathcal{B}$, first we need to construct $\mathcal{BB}$ based on the images and their features in the experimental database. Based on $\mathcal{BB}$, $\mathcal{B}$ can be obtained using the procedure aforementioned. The initial state probability distribution for experimental database can be determined by using Equation (5). The constructions of these model parameters can be performed off-line.

Table 5. Image retrieval steps using the stochastic model

1. Given the query image q, obtain its feature vector $query = \{o_1, o_2, \ldots, o_T\}$, where T is the total number of non-zero features of the query image q.
2. Upon the first feature o_1, calculate $W_1(i, j)$ and $D_1(i, j)$ according to Equations (6) and (7).
3. Move on to calculate $W_2(i, j)$ and $D_2(i, j)$ according to Equations (8) and (9).
4. Continue to calculate the next values for the W and D matrices until all the features in the query have been taken care of.
5. Upon each feature in query, we can obtain a pair of matrices: $W_t(i, j)$ and $D_t(i, j)$. We then sum up each column in matrices $W_t(i, j)$ and $D_t(i, j)$. Namely, we calculate $sumW_t(j) = \sum_i W_t(i, j)$ and $sumD_t(j) = \sum_i D_t(i, j)$.
6. Find the candidate images by sorting their corresponding values in $D_t(q, j)$, $sumD_T(j)$, $sumD_{T-1}(j)$, …, or $sumD_1(j)$. First, an image is ranked according to its value in $D_T(q, j)$. If there exist several images that have the same value, then $sumD_T(j)$ values are used for ranking. If several images have the same $sumD_T(j)$ value, then $sumD_{T-1}(j)$ values are used and the process continues until $sumD_1(j)$.
7. Select the top ranked images from the output of Step 6, and rank them to the user based on their values in $W_T(q, j) + W_T(j, q)$.

Once the model parameters of the MMM for the image database is constructed, the stochastic process shown in Table 5 is used for image retrieval.

Stochastic Process for Example Queries

For a given query image issued by a user, the stochastic process with the proposed dynamic programming algorithm will be carried on to dynamically find the matched images for the

user's query. A series of W_t and D_t matrices are generated according to Equations (6) to (9). The qualifying degrees of the images with respect to the certain query image are estimated by the values in the resulting W_t and D_t matrices according to rules described in Table 5.

In this section, we use a set of example queries to demonstrate the effectiveness of our stochastic model. For each set of query results, the qualifying possibilities of the images are in the descending order from the *top left* to the *bottom right*. The searching results are listed and analyzed as well. As can be seen from the experimental results, our method effectively extracts the images that contain the features specified in the query image and ranks them appropriately.

Query I

In this query example, the query image #2265 has one color feature and one location feature which are 'blue' (b) and 'L5.' Exhibit 1 gives the corresponding $\mathcal{B}$ matrix entry for it.

Since there are two features in the query image, two *W* matrices and two *D* matrices will be generated. The system is supposed to return those images that have the desired features in the ordering of similarity. The snapshot of the retrieval result screen containing the top 12 images is shown in Figure 3. As can be seen from this figure, the 'blue' color is the dominating color in all the retrieved images. Moreover, they all have an object located at 'L5' (the centre location). It should also be noted that some images are excluded from the query results because their values in $sumW_T(j)$ are zeros, which means they do not have the corresponding features ('blue' or 'L5').

Query II

In this query, the major features of this query image include four components: 'blue' (b), 'white' (w), 'yellow-green' (yg), and 'L5' (the centre location) with the $\mathcal{B}$ matrix in Exhibit 2.

The most qualified images to this query are those that have all of the above three-color features and an object at the L5 location (i.e., the centre location). Images that have only one of the desired features are less satisfactory. The snapshot of this example query is given in Figure 4. All the retrieved top 12 images have the above mentioned color features and have the object(s) at location 'L5.'

Exhibit 1.

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img2265	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0

Figure 3. Snapshot of Query I

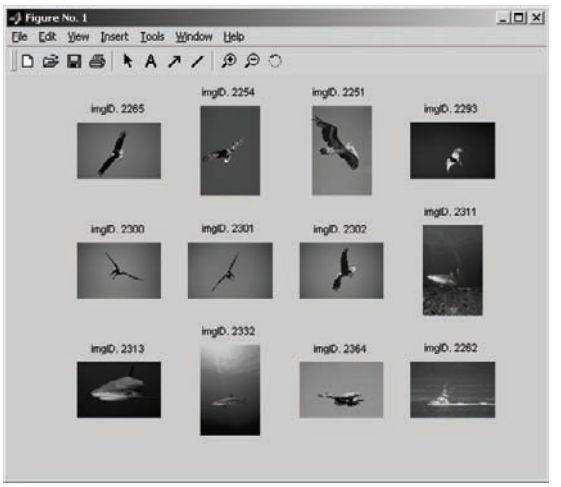


Exhibit 2.

black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
0	0.17	0	0	0	0.17	0	0	0.16	0	0	0	0	0	0	0	0.5	0	0	0	0

Query III

Similar to the previous query example, there are four features, which are 'red,' 'purple-red,' 'L1,' and 'L5.' Two of them are the color features and the other two are the object location descriptions. Figure 5 exhibits the top 24 images retrieved from the image database. The final query results are good and the ranking is reasonable.

FUTURE TRENDS

The current CBIR systems are promising and reflect the increasing interest in the field of content-based image retrieval. However, there are still a number of open research issues to be addressed. For example, it is critical to develop the suitable evaluation criteria and benchmarks for CBIR systems. Some future trends include:

- *Automatic or semi-automatic methods of object extraction for image retrieval:* It has been recognized that the searching for images in large databases can be greatly facilitated by the use of semantic information such as object location and type. However, using the current technique of computer vision cannot extract the object information easily. For example, even though the unsupervised image segmentation is applied in our framework to obtain the object informa-

Figure 4. Snapshot of Query II

Figure 5. Snapshot of Query III

tion, we still have to find a trade-off between the accuracy and performance. Considering that users may upload their own query image (which is not in the database) during the content-based retrieval, the real-time feature extraction could be a big issue to affect the users' will to use the system. There is an increasing need to derive more efficient yet good enough methods of segmenting images to distinguish objects of interest from their background. The main goal is to bridge the semantic gap, bringing some measure of automation to the processes of indexing and retrieving images by the type of object or scene depicted.

- *Indexing standard and query language support for image databases:* Currently, there are neither the standards in indexing the features and the subject taxonomies for image databases, nor the standard for hierarchical representation of an image when taking into consideration the objects inside the image. These standards are essential in expressing complex semantics and supporting the manipulation of content-based queries on image objects. Once the feature representations of image objects follow the same standards, it becomes possible to develop a suitable query language exclusively designed for image databases.
- *Support for automatic video indexing and retrieval:* Recently, lots of research work has been done on automatic video segmentation. After video is segmented into smaller units such as shots or scenes, each unit is represented by its key frames. Then, the complex spatial-temporal semantics can be obtained by parsing the content of these key frames, which forms the basis for supporting advanced video retrieval techniques such as query-by-motion facilities. MPEG-7 (Moving Picture Experts Group) standard (*http://media.wiley.com/product_data/excerpt/87/04714867/0471486787.pdf*) is the first to take into consideration the issue of multimedia content representation. This standard will define a standard for describing every aspect of the content of a multimedia object, including the specifications of a video's image features. MPEG-7 will definitely have an impact on CBIR and will probably guide the development of future CBIR systems.
- *Better user interaction, especially the improved techniques for collecting users' feedback:* User's relevance feedback has been adopted in most recent efforts towards the research of CBIR. In order to get better results, the user may be asked to browse a bunch of images through iterations and to provide the detailed ranking for similarity for the images. The fact is that a heavy and unnecessary burden of responsibility is brought to the user. In addition, it is highly probable that this burden will have a negative effect on a user's perception of the effectiveness and efficiency of the system.

CONCLUSION

Currently, Content-Based Image Retrieval (CBIR) technology is still immature but with great potential. In this chapter, a review of the recent efforts and techniques in CBIR is given, followed by the discussion of the current problems in the CBIR systems from the efficiency concern of the searching process. In response to this issue, in this chapter, the Markov model mediator (MMM) mechanism is applied to the image databases for content-based image retrieval. A stochastic process based on the MMM mechanism is proposed to traverse the database and find the similar images with respect to the query image. This approach performs similarity comparison based on not only the relationship between the query image and the target image, but also the relationships among all the images within the database, such as their access patterns and access frequencies. Joint color/layout similarity measurement is supported in this system to offer more complete distinction descriptions of the images and better retrieval effectiveness. Several experiments were conducted and the experimental query-by-example image query results to the proposed retrieval system were reported. The fact that the proposed stochastic content-based CBIR system utilizes the MMM mechanism and supports both spatial and color information offers more flexible and accurate results for user queries. The experimental results exemplify this point and the overall retrieval performance of the presented system is promising. Besides modelling the relationship of

image objects within a single database, the MMM mechanism also has the capability (although not shown in this chapter) to model the relationships among distributed image databases so as to guide the efficient search across the distributed image databases.

ACKNOWLEDGMENT

For Shu-Ching Chen, this research was supported in part by NSF EIA-0220562 and NSF CDA-9711582.

REFERENCES

Androutsos, D. (1999). *Efficient indexing and retrieval of color image data using a vector-based approach*. Ph.D. Dissertation. Department of Electrical and Computer Engineering, University of Toronto.

Aslandogan, Y. A., & Yu, C. T. (1999). Techniques and systems for image and video retrieval. *IEEE Trans. On Knowledge and Data Engineering, 11*(1), 56-63.

Carson, C. S., Belongie, S., Greenspan, H., & Malik, J. (1997). Region-based image querying. In *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, (pp. 42-49).

Chen, S.-C., Sista, S., Shyu, M.-L., & Kashyap, R.L. (2000). An indexing and searching structure for multimedia database systems. *IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2000,* (pp. 262-270).

Cheng, H.D. & Sun, Y. (2000). A hierarchical approach to color image segmentation using homogeneity. *IEEE Transactions on Image Processing, 9*(12), 2071-2082.

Cheng, H.D., Jiang, X.H., Sun, Y., & Wang, J.L. (2001). Color image segmentation: Advances and prospects. *Pattern Recognition,* 34, 2259-2281.

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic D., & Equitz, W. (1994). Efficient and effective querying by image content. *Journal of Intelligent Information Systems, 3*(3/4), 231-262.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by image and video content: The QBIC system. *IEEE Computer, 28*(9), 23-31.

Frank, O. & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81,* 832-842.

Guttman, A. (1984). R-tree: A dynamic index structure for spatial search. *Proc. ACM SIGMOD,* (pp. 47-57).

Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M., & Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*(7), 729-736.

Kaplan, L. M., Murenzi, R., & Namuduri, K. R. (1998). Fast texture database retrieval using extended fractal features. In *Storage and Retrieval for Image and Video Databases VI, Proc. SPIE, 3312,* (pp. 162-173).

Li, Z.-N., Tauber, Z., & Drew, M. S. (2000). Locale-based object search under illumination change using chromaticity voting and elastic correlation. In *Proc. IEEE International Conference on Multimedia and Expo (II) 2000,* (pp. 661-664).

Lin, H. C., Wang L. L., & Yang, S. N. (1997). Color image retrieval based on hidden Markov models. *IEEE Transactions on Image Processing, 6*(2), 332-339.

Ma, W.-Y. & Manjunath, B. S. (1998). A texture thesaurus for browsing large aerial photographs. *Journal of the American Society for Information Science*, *49*(7), 633-648.

Ma, W.-Y. & Zhang, H.J. (1999). Content-based image indexing and retrieval. *Handbook of Multimedia Computing*, CRC Press.

Naphade, M. R. & Huang, T. S. (2001). A probabilistic framework for semantic indexing and retrieval in video. *IEEE Transactions on Multimedia*, *3*(1), 141-151.

Pentland, A., Picard, R.W., & Sclaroff, S. (1994). Photobook: Tools for content-based manipulation of image databases. In *Proc. Storage and Retrieval for Image and Video Databases II, 2185*, (pp. 34-47), SPIE.

Rabiner, L. R. & Huang, B. H. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, *3*(1), 4-16.

Rui, Y. & Huang, T. S. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, *10*(4), 39-62.

Rui, Y., Huang, T. S., & Mehrotra, S. (1997). Content-based image retrieval with relevance feedback in MARS. In *Proceedings of the 1997 International Conference on Image Processing (ICIP'97)*, (pp. 815-818).

Shyu, M.-L., Chen, S.-C., & Haruechaiyasak, C. (2001a). Mining user access behavior on the www. *IEEE International Conference on Systems, Man, and Cybernetics*, (pp. 1717-1722).

Shyu, M.-L., Chen, S.-C., & Kashyap, R. L. (2000a). A probabilistic-based mechanism for video database management systems. *IEEE International Conference on Multimedia and Expo (ICME2000)*, (pp. 467-470).

Shyu, M.-L., Chen, S.-C., & Kashyap, R. L. (2000c). Organizing a network of databases using probabilistic reasoning. *IEEE International Conference on Systems, Man, and Cybernetics*, 1990-1995.

Shyu, M.-L., Chen, S.-C., & Shu, C.-M. (2000b). Affinity-based probabilistic reasoning and document clustering on the www. *The 24th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC)*, (pp. 149-154).

Shyu, M.-L., Chen, S.-C., Haruechaiyasak, C., Shu, C.-M., & Li, S.-T. (2001b). Disjoint web document clustering and management in electronic commerce. *The Seventh International Conference on Distributed Multimedia Systems (DMS'2001)*, (pp. 494-497).

Sista, S. & Kashyap, R. L. (1999). Unsupervised video segmentation and object tracking. *IEEE International Conference on Image Processing.*

Smith, J. R. & Chang, S. F. (1996). VisualSEEK: A fully automated content-based image query system. In *Proceedings ACM International Conf. Multimedia*, (pp. 87-98).

Stricker, M. & Dimai, A. (1996). Color indexing with weak spatial constraints. In *Storage and Retrieval for Image and Video Databases IV, Proc SPIE 2670*, (pp. 29-40).

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computers*, *25*(3), 38-49.

Wolf, W. (1997). Hidden Markov model parsing of video programs. Presented at the *International Conference of Acoustics, Speech and Signal Processing.*

Zhong, Y., Jain, A.K., & Dubuisson-Jolly, M.-P. (2000). Object tracking using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *22*(5), 544-549.

This work was previously published in Multimedia Systems and Content-Based Image Retrieval, edited by S. Deb, pp. 302-321, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.12
A Spatial Relationship Method Supports Image Indexing and Similarity Retrieval

Ying-Hong Wang
Tamkang University, Taiwan

ABSTRACT

The increasing availability of image and multimedia-oriented applications markedly impacts image/multimedia file and database systems. Image data are not well-defined keywords such as traditional text data used in searching and retrieving functions. Consequently, various indexing and retrieving methodologies must be defined based on the characteristics of image data. Spatial relationships represent an important feature of objects (called icons) in an image (or picture). Spatial representation by 2-D String and its variants, in a pictorial spatial database, has been attracting growing interest. However, most 2-D Strings represent spatial information by cutting the icons out of an image and associating them with many spatial operators. The similarity retrievals by 2-D Strings require massive geometric computation and focus only on those database images that have all the icons and spatial relationships of the query image. This study proposes a new spatial-relationship representation model called "Two Dimension Begin-End boundary string" (2D Bε-string). The 2D Bε-string represents an icon by its MBR boundaries. By applying "dummy objects," the 2D Bε-string can intuitively and naturally represent the pictorial spatial information without any spatial operator. A method of evaluating image similarities, based on the modified "Longest Common Subsequence" (LCS) algorithm, is presented. The proposed evaluation method can not only sift out those images of which all icons and their spatial relationships fully accord with query images, but for those images whose icons and/or spatial relationships are similar to those of query images. Problems of uncertainty in the query targets and/or spatial relationships thus solved. The representation model and similarity evaluation also simplify the retrieval progress of linear transformations, including rotation and reflection, of images.

INTRODUCTION

The increasing availability of image and multimedia-oriented applications very significantly influences image/multimedia file and database systems. Image data are not well-defined keywords such as traditional text data used in searching and retrieving functions. Therefore, various indexing and retrieving methodologies have to be defined based on the characteristics of image data.

Abstracting the information in original images is very important in pictorial spatial application systems. These abstractions include how the image icons and their characteristics were recognized; how the symbolic image was encoded and constructed; how to index and retrieve these images; how to evaluate their similarity corresponding to a query image, and many others. All of these are very important issues in information and content-based retrieval.

Three basic types for image indexing and retrieval exist: (1) by features, for example, color, texture, or shape, of the icons in the images, as in the QBIC project (Huang & Jean, 1994), and the Virage search engine (Liang & Mou, 1996); (2) by size and location of the image icons, as in R-tree (Guttman, 1984), R*-tree (Beckmann, Kriegal, Schneider, & Seeger, 1990), Quadtree (Samet, 1989), and Mou's similarity method (Chien, 1998); (3) by relative position of the icons (called Spatial relationship), as in 2-D String (Chang, Shi, & Yan, 1987) and its variants (Chang, Jungert, & Li, 1988; Lee & Hsu, 1990; Lee, Yang, & Chen, 1992; Bach, et al., 1996; Liang & Mou, 1997; Hsu, Lee, & Lin, 1999; Sipser, 1997; Papadias & Theodoridis, 1995). Spatial relationships represent an important feature of objects (called icons) in an image (or picture).

The third type is most suited to those applications that are independent of the actual coordinates of icon objects, as in for example, 'find all images in which icon *A* is located on the left side and icon *B* on the right.' Spatial representation by 2-D String and its variants, in a pictorial spatial database, has attracted increasing attention. However, most of 2-D Strings represent the spatial information by cutting the icons out of an image, which are associated with many spatial operators. Their similarity retrievals require extensive geometric computation and focus only on those database images that include all of the icons and spatial relationships of the query image.

This investigation proposes a novel spatial knowledge representation model called, "Two Dimension Begin-End boundary string (2D Bε-string)." The 2D Bε-string need not cut any image's icons because it directly represents an icon by the boundaries of its MBR (Minimum Bounding Rectangle). By applying "**dummy** objects," the 2D Bε-string can intuitively and naturally represent the pictorial spatial information without any spatial operators. An algorithm is also introduced, which involves $\boldsymbol{O(n)}$ space complexity in the best and worst case, to establish an image database using the proposed model.

A method of evaluating image similarities is also presented, according to the modified "Longest Common Subsequence" (LCS) algorithm (Cormen, Leiserson, & Rivest, 1990). The proposed evaluation method sifts out not only those images in which all of the icons and their spatial relationships fully accord with those of the query images, but also those images whose icons and/or spatial relationships are similar to those of the query image. Problems of uncertainty of the query targets and/or spatial relationships are thus solved. The modified LCS algorithm involves $\boldsymbol{O(mn)}$ space and time complexity, where m and n are the numbers of icons in a query image and a database image, respectively. Retrieving the linear transformations of an image represented by a 2D Bε-string is simpler than other models of 2-D Strings. The transformations include 90, 180, 270 degrees clockwise rotations and reflections in the x- and y-axis.

The rest of this chapter is organized as follows. The chapter reviews the approaches of 2-D String and its variants and is followed with a section that

proposes a new spatial knowledge representation model, called 2D Bε-string, and an algorithm to construct a symbolic picture using 2D Bε-string. Then, the chapter introduces a similarity retrieval algorithm, modified from LCS, and the corresponding similarity evaluation process. Next, the chapter presents a demonstration system, which is a visual retrieval system, implemented by 2D Bε-string and the modified LCS algorithm. The chapter finishes with concluding remarks and suggests directions for future work.

BACKGROUND

This section introduces relevant background theories and techniques.

Minimum Bounding Rectangle (MBR)

MBR is the most popular approximation to identify an object from images (Papadias & Theodoridis, 1995; Bowen & Um, 1999; Song, Whang, Lee, & Kim, 1999; Orlandic & Yu, 2000), and retains the most important characteristics of the object; position and extension. MBR is characterizes an object by minimum and maximum *x* and *y* coordinates (Zimbrao & De Souza, 1998). Figure 1 presents a coordinate system for a picture/image considered in this chapter. Symbols minX(*A*) and maxX(*A*) are minimum and maximum *x* coordinates of the MBR for an object, *A*. Symbols minY(*A*) and maxY(*A*) are minimum and maximum *y* coordinates of the MBR for an object, *A*.

Representation Models of 2-D String and its Variants

Chang, Shi and Yan (1987) proposed an approach, called '2-D String', to represent the spatial information of a picture or image. The 2-D String uses a symbolic projection of a picture on the *x*- and *y*-axes. They define two sets of symbols, *V* and *A*. Each symbol in *V* presents an icon object in an image. *A* is a set of spatial operators containing {'=', '<', ':'}. A 2-D String over *V* and *A* is defined by $(u, v) = (o_1x_1o_2x_2\ldots x_{n-1}o_n, o_{p(1)}y_1o_{p(2)}y_2\ldots y_{n-1}o_{p(n)})$, where $o_1o_2\ldots o_n$ is a 1-D string over *V*, $x_1x_2\ldots x_{n-1}$ and $y_1y_2\ldots y_{n-1}$ are 1-D strings over *A*; $o_{p(1)}o_{p(2)}\ldots o_{p(n)}$ is a permutation of $o_1o_2\ldots o_n$ on the *y*-axis. Therefore, *n* icon objects are in the picture/image and their spatial information is represented by operators, $x_1x_2\ldots x_{n-1}$ and $y_1y_2\ldots y_{n-1}$ on *x*-axis and *y*-axis, respectively. Figure 2 depicts an example of 2-D String.

The 2D G-string (Chang, Jungert, & Li, 1988), a variant of 2-D String, extends spatial relationships to two sets of spatial operators, R_l and R_g, and cuts all the objects along their MBR boundaries. The set, R_l, defines local spatial relationships that are partially overlapping projections between two objects. The set, R_g, defines global spatial relationships, such that the projection of two objects are disjoining, adjoining or in the same position.

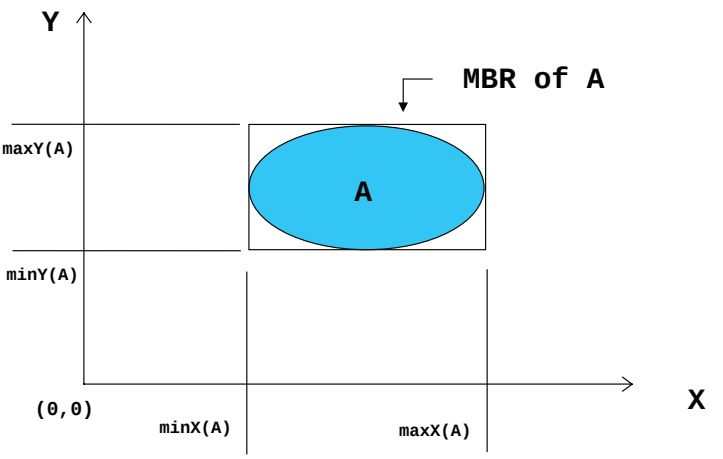
The 2D C-string (Lee & Hsu, 1990; Bach, et al., 1996), another variant of 2-D String, minimizes the number of cutting objects. The 2D C-string keeps the leading object uncut and addresses the problem of superfluous cutting objects generated by 2D G-string cutting progress. The worst case will involve $O(n^2)$ cutting objects.

Consequently, the 2D B-string (Lee, Yang, & Chen, 1992) does not use cutting, and instead, represents an object by two symbols. One of which indicates for the beginning boundary of that object; the other one represents the end. 2D B-string also reduces the spatial relationships to a single operator '=', such that two objects have the same boundary projection if '=' appears.

Similarity Retrieval and Evaluation

The basic similarity retrieval and evaluation approaches of 2-D String (Chang, Shi, & Yan, 1987), 2D G-string (Chang, Jungert, & Li, 1988), 2D C-string (Lee & Hsu, 1990) and 2D B-string (Lee, Yang, & Chen, 1992) are the same (Chan & Chang, n.d.). First, they always define three types

Figure 1. MBR presentation of an object in a picture/image



of similarity, type-*i* (*i* = 0, 1, 2). Each is constrained according to some conditions. Type-1 is stricter than type-0; type-2 is stricter than type-1. Second, they compare all spatial relationship pairs between any two objects in the query image to pairs in an image in databases and build a type-*i* subgraph if the pair satisfies type-*i* constraints. Then, they find the **maximum complete subgraph** from each existing type-*i* graph. The number of objects in the maximum complete subgraph measures the similarity of the query image to images in databases.

The space and time complexity of examining all spatial relationship pairs is $O(n^2)$, where n is the number of objects in an image. Finding the maximum complete subgraph is an NP-complete problem (Kim & Um, 1999), and is a time consuming procedure is not suited to a large number of icon objects in an image.

SPATIAL REPRESENTATION MODEL USING 2D Bε-STRING

2D Bε-string Model

Several approaches have been proposed to represent an icon in an image. These include MBR (Minimum Bounding Rectangle) (Chang, Jungert, & Li, 1988; Lee & Hsu, 1990; Lee, Yang, & Chen, 1992; Bach, et al., 1996), MBE (Minimum Bounding Ellipse), MBC (Minimum Bounding Circle) (Chien, 1998), and others.

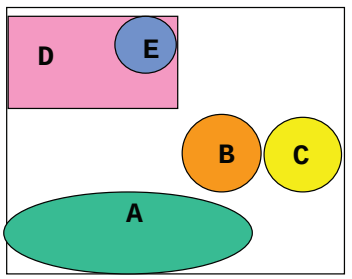
The approach presented here, 2D Bε-string, is based on MBR. Conceptually, it is similar to the 2D B-string approach. Both represent an object by its MBR boundaries and require nothing to be cut. However, 2D Bε-string states the spatial relationship between two boundary symbols in quite a different way. 2D B-string uses a **spatial operator** (=) to denote an **IDENTITY** between the projections of two boundaries. In 2D Bε-string, a **dummy object** is used to describe the **DISTINCTION** between the projections of two boundaries.

The definition of 'Dummy Object' is as follows:

A 'Dummy Object' is an assumed virtual object; it is not a real object existing in an original image. It can be specified as any size of space and be memorized as symbol 'ε'.

2D Bε-string can be defined as:

Figure 2. Example of 2D string representation



The 2D String is

(A=D:E<A=B<C, A<B=C<D:E)

$(u, v) = (d_0x_1d_1x_2d_2\ldots d_{n-1}x_nd_n,\ d_0y_1d_1y_2d_2\ldots d_{n-1}y_nd_n)$

where d_i is a dummy object ε, or a null string, $i = 0, 1, \ldots n$, and x_i and y_i are real icon objects which are either beginning or ending projected boundaries on the *x*- and *y*-axis, respectively, with $i = 1, 2, \ldots n$. d_i can be determined when the maximum size of an image, say X_{max} and Y_{max} along the *x*- and *y*-axes, respectively, is known. Set d_0 to ε if a space exists between the beginning boundary of the leftmost (bottommost) object and the left (bottom) edge of the image. Similarly, set d_n to ε if an interval exists between the ending boundary of the rightmost (topmost) object and the right (top) edge of the image. For the dummy objects, set d_i to ε if the boundary projections of x_i and x_{i+1} (y_i and y_{i+1}) differ from each other.

The 2D Bε-string in Figure 3, for example, is written as $(u, v) = (\varepsilon A_b \varepsilon B_b \varepsilon A_e C_b \varepsilon C_e \varepsilon B_e \varepsilon, \varepsilon B_b \varepsilon A_b \varepsilon B_e C_b \varepsilon C_e \varepsilon A_e \varepsilon)$. The dummy object, d_0, was set to ε because a space exists in front of the beginning boundary of object *A* on the *x*-axis. The dummy object d_6 was also set to ε because a space is in the behind object, *B*, on the *x*-axis. However, the dummy object, d_3, was set to null string, because the ending boundary of object *A* and the beginning boundary of object *C* are projected onto the same point on the *x*-axis. A similar case applies to the ending boundary of object *B* and the beginning boundary of object *C* on the *y*-axis.

Observably, 2D Bε-string has the following advantages:

First, the object location in the original image and the symbolic picture are mapped directly, as show in Table 1. No operator is required to represent the spatial relationship between objects. The approach is intuitive.

Second, this approach of 2D Bε-string does not need to cut the objects from the image. It simplifies the construction of the image database. The space complexity for an image with *n* objects in the worst and best case is *O*(n). In the worst case, all boundary projections are distinct and a space is left in the leftmost, bottommost, rightmost and topmost parts of an image ($4n+1$ symbols are required). In contrast, in the best case, all boundary projections are identical and exactly fit in an image (requiring $2n+1$ symbols only).

Third, the approach simplifies similarity retrieval because results of spatial reasoning need not be combined.

Algorithm for Constructing a Symbolic Picture

By default, before converting an image to a symbolic picture represented by 2D Bε-string, all objects and their MBR coordinates must be abstracted from that image. Then, algorithm **Convert-2D-Bε-String**, shown in Table 2, is called to transform an original image into a symbolic picture. Lines one to 12 explain the meaning of the variables used in the input parameters and the conversion process. Lines 14 to 19 separately sort the input data by key coordinate and object identifiers, in ascending order on the *x*- and *y*-axes. Lines 21 to 32 construct the 2D Bε-string on the *x*-axis, and lines 34 to 45 construct the 2D Bε-string on the *y*-axis.

The time complexity of algorithm **Convert-2D-Bε-String** depends on the sorting algorithm in line 19. The time complexity of loops in lines 14 to 18, 24 to 30 and 37 to 43 is $O(n)$, and never exceeds that of the sorting algorithm. Ignoring the sorting algorithm, the space complexity of this algorithm is also $O(n)$.

Building an image database using 2D Bε-string, requires only algorithm **Convert-2D-Bε-String** to be called with the MBR coordinates and object identifiers of each image and the results of the 2D Bε-string to be saved in the database. Saving the 2D Bε-string with the MBR coordinates of objects in an image allows the location into which a new object is to be inserted and the MBR boundaries of that object to be easily found using the binary search method with MBR coordinates and the identifier of the new object as searching keys. This is because the 2D Bε-string uses an order datum to represent the relationship among the objects in an image. Whether a dummy object is to be inserted around this new object boundary is easy to determine. In contrast with insertion, dropping an object from an image, requires the dropping object to be searched sequentially from the 2D Bε-string data. It is a direct deletion and the redundant dummy object, if found, to be deleted.

Figure 3. Image with three objects

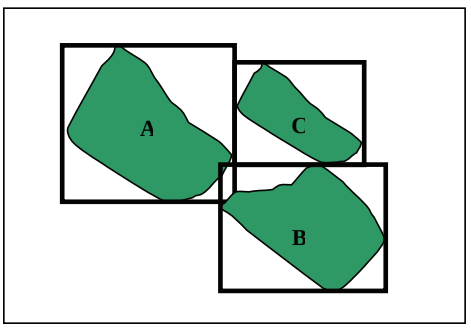


IMAGE SIMILARITY RETRIEVAL AND EVALUATION

Assessment of Similarity

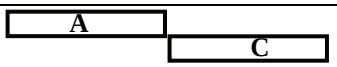
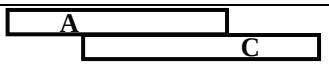
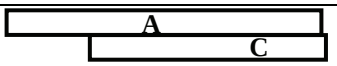
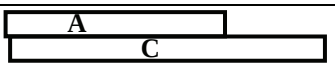
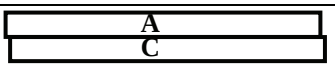
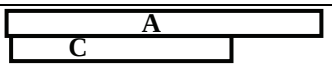
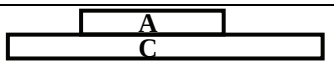
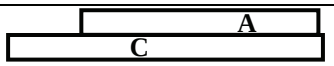
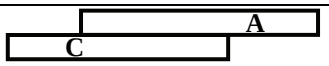
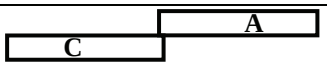
The "Background" section discussed the symbolic image construction and the assessment of similarity, using 2D B-String. In order to decide which type, type-0, type-1 or type-2, of similarity it belongs to all spatial relationships for every pair of objects in a database image must be recorded to determine of which type is the similarity. However, this assessment does not address a situation in which only a few query objects are present in a database image. Moreover, the assessment of similarity as type-0, type-1 and type-2 is based on the maximum complete subgraph. Obtaining a maximum complete subgraph from C_2^m spatial relationships of m objects in a query image and C_2^n spatial relationships of n objects in the database image is rather time-consuming.

This chapter also uses the number of spatial relationships, formed by every pair of objects that simultaneously appear in the query image and the database image, as the assessment of similarity. However, the proposed method does not find a maximum complete subgraph but finds a longest common subsequence (LCS) (Cormen, Leiserson,

& Rivest, 1990) of two 2D Bε-strings instead. After the LCS string is found, it is evaluated in relation to 2D Bε-strings of the query image and the database image.

The motivations for finding an LCS string in this work and finding a maximum complete subgraph in 2-D Strings from two images are different tunes rendered with equal skill. Both the LCS string and the subgraph are used to evaluate similarity. However, the time complexity of an LCS algorithm is $O(mn)$, where m and n are the number of objects in the query image and the database image, respectively. The complexity depends only on the length of 2D Bε-string in the query image and the database image. Not all the spatial relationships need be examined for every pair of objects because **"in the query image and the database image, all spatial relationships of every pair of objects in the LCS string are identical."** Therefore, similarity can be evaluated in a reasonable time.

Table 1. Mapping of spatial relationships

	In symbolic picture	In original image
1	$\varepsilon A_b \varepsilon A_e \varepsilon C_b \varepsilon C_e \varepsilon$	
2	$\varepsilon A_b \varepsilon A_e C_b \varepsilon C_e \varepsilon$	
3	$\varepsilon A_b \varepsilon C_b \varepsilon A_e \varepsilon C_e \varepsilon$	
4	$\varepsilon A_b \varepsilon C_b \varepsilon C_e A_e \varepsilon$	
5	$\varepsilon A_b \varepsilon C_b \varepsilon C_e \varepsilon A_e \varepsilon$	
6	$\varepsilon A_b C_b \varepsilon A_e \varepsilon C_e \varepsilon$	
7	$\varepsilon A_b C_b \varepsilon A_e C_e \varepsilon$	
8	$\varepsilon A_b C_b \varepsilon C_e \varepsilon A_e \varepsilon$	
9	$\varepsilon C_b \varepsilon A_b \varepsilon A_e \varepsilon C_e \varepsilon$	
10	$\varepsilon C_b \varepsilon A_b \varepsilon A_e C_e \varepsilon$	
11	$\varepsilon C_b \varepsilon A_b \varepsilon C_e \varepsilon A_e \varepsilon$	
12	$\varepsilon C_b \varepsilon C_e A_b \varepsilon A_e \varepsilon$	
13	$\varepsilon C_b \varepsilon C_e \varepsilon A_b \varepsilon A_e \varepsilon$	

Algorithms of Similarity Retrieval

This study presents an algorithm to find the longest common subsequence length of two 2D Bε-strings. Table 3 shows this algorithm, which is called, **2D-Bε-LCS-Length**. The algorithm is modified from the LCS algorithm in Cormen, Leiserson and Rivest (1990). Revising the original LCS algorithm depends on two considerations. First, LCS is prevented from picking dummy objects continuously, because a single dummy object sufficiently represents the relative spatial relationship between two boundary symbols. The *if*-statement in line 21 performs this evaluation. Second, the LCS paths-recording matrix in the original LCS algorithm is omitted by evaluating left and up paths first, as shown in lines 16 to 19, and then evaluating the left-up diagonal path, as shown lines 23 to 24. The LCS path can still be inferred from the matrix record of the LCS length.

Algorithm **2D-Bε-LCS-Length** uses two 2D Bε-strings, *Q* and *D*, as input parameters — one for the query image and the other one for the database image. The 2D Bε-string of a query image with *m* objects in, each dimension has 2*m* boundary symbols and up to 2*m*+1 dummy objects. The maximum length of a 2D Bε-string in each dimension is 2*m*+(2*m*+1), that is, 4*m*+1. The maximum length of 2D Bε-string of a database image with *n* objects is similarly derived as 4*n*+1. The LCS length-inferring table *W*, requires (1+(4*m*+1))(1+(4*n*+1)) storage units; therefore, the space complexity is *O*(*mn*).

Lines seven to eight and 10 to 11 of the algorithm in Table 2 show the initialization of the first row and first column in the LCS-length inferring table, *W*: each string symbol must be set once. These instructions are executed 4*m*+1 and 4*n*+2 times, respectively. The outer loop, in line 13, examines each row in *W*, 4*m*+1 times. The inner loop, in lines 14 to 26, examines each cell of *W*, (4*m*+1)*(4*n*+1) times. Thus, the time complexity is *O*((4*m*+1)+ (4*n*+2)+(4*m*+1)*(4*n*+1)) the same as *O*(*mn*).

A recursive procedure is also specified for printing the longest common subsequence string of two 2D Bε-strings. Table 4 shows this algorithm. The initial invocation is **Print-2D-Bε-LCS** (*Q*, *W*, *length*(*Q*), *length*(*D*)). From the value of the last cell in the LCS-inferring table, *W*, this procedure decreases by *i* and/or *j* in the left and/or up direction in each recursive call, until either *i* or *j* reaches zero. Then, all symbols of the LCS string are printed out in the proper, forward order.

To find the longest common subsequence length of two 2D Bε-strings, the LCS string length of the current cell must be compared with that of the top one before the LCS string symbol is printed, according the revised procedure shown as Table 3. Therefore, the LCS path is determined from the up direction if the LCS string lengths are equal. The corresponding boundary symbol or dummy object is not a symbol of the LCS string. This symbol is thus ignored and is induced continuously along the up direction, as shown in lines five to six. However, if the lengths are unequal, the current cell's LCS string length must be compared with the left cell's length. If these lengths are the same, then the LCS is induced from the left direction. For the same reason as in the preceding case, if the corresponding boundary symbol or dummy object is not a symbol of the LCS string, this symbol is ignored and is induced continuously along the left direction, as shown in lines seven through eight. If the LCS is from neither the up nor the left direction, it must be induced from the left-up diagonal direction. The boundary symbol or dummy object associated with the current cell must be part of the LCS string. After recursively processing all the cells in the left/up direction, this symbol in the LCS string is printed out, as shown in lines nine to 11.

Lines six, eight and 10 are recursively called; each time, either *i* or *j* is reduced by one. This algorithm can print out all LCS string symbols after no more than *m*+*n* recursions. The time complexity is *O*(*m*+*n*).

Similarity Evaluation

After the LCS length and the LCS string are obtained from the query image and data images in each dimension of 2D Bε-string, how is similarity evaluated? Conceptually, the two images are more similar if the LCS string is longer. For example, the strings in one dimension of 2D Bε-string are:

Query image 1: $\varepsilon B_b \varepsilon C_b \varepsilon B_e \varepsilon A_b \varepsilon C_e \varepsilon A_e \varepsilon$,
Database image 1: $\varepsilon B_b \varepsilon B_e \varepsilon D_b \varepsilon A_b \varepsilon D_e \varepsilon A_e \varepsilon$,
LCS string 1: $\varepsilon B_b \varepsilon B_e \varepsilon A_b \varepsilon A_e \varepsilon$,
Database image 2: $\varepsilon B_b \varepsilon C_b \varepsilon B_e \varepsilon D_b \varepsilon A_b \varepsilon D_e \varepsilon C_e \varepsilon A_e \varepsilon$,
LCS string 2: $\varepsilon B_b \varepsilon C_b \varepsilon B_e \varepsilon A_b \varepsilon C_e \varepsilon A_e \varepsilon$.

The LCS strings 1 and 2 follow from query image 1 with database images 1 and 2, respectively, according to the LCS-length-inferring algorithm in Table 3 and the LCS path-discovering algorithm in Table 4. For a given query image string, database image 2 yields a better result than database image 1 because LCS string 2 is longer than LCS string 1. However, two database images with the same LCS string length may not have the same degree of similarity. For example:

Query image 2: $\varepsilon B_b \varepsilon C_b \varepsilon B_e A_b \varepsilon C_e \varepsilon A_e$,
Database image 3: $\varepsilon B_b \varepsilon B_e \varepsilon D_b \varepsilon A_b \varepsilon D_e \varepsilon A_e \varepsilon$,
LCS string 3: $\varepsilon B_b \varepsilon B_e A_b \varepsilon A_e$,
Database image 4: $\varepsilon B_b \varepsilon D_b \varepsilon B_e A_b \varepsilon D_e \varepsilon A_e \varepsilon$,
LCS string 4: $\varepsilon B_b \varepsilon B_e A_b \varepsilon A_e$,

The length of LCS strings 3 and 4 is the same, but database image 4 gives a better matching result than database image 3, because query image 2 requires object *B* and object *A* to be adjoined, that is, boundary symbols B_e and A_b must have the same projection. But in database image 3, object *A* is not adjoined to object *B* because spaces exist between boundary symbols B_e and A_b. The interval does not exist in database image 4. Accordingly, database image 4 has a better similarity than database image 3.

Following the above analysis, the similarity of preceding phenomena is assessed. The following notation is first defined over 2D Bε-string:

Table 2. Algorithm to construct a symbolic picture

Convert-2D-Bε -String (n, C, X_b, X_e, Y_b, Y_e, x_{max}, y_{max})

1. / / n …number of objects in an image
2. / / C …object symbols in image, $C = \{c_1, c_2, \ldots, c_n\}$
3. / / X_b …begin boundaries on x-axis, $X_b = \{x_{b1}, x_{b2}, \ldots, x_{bn}\}$
4. / / X_e …end boundaries on x-axis, $X_e = \{x_{e1}, x_{e2}, \ldots, x_{en}\}$
5. / / Y_b …begin boundaries on y-axis, $Y_b = \{y_{b1}, y_{b2}, \ldots, y_{bn}\}$
6. / / Y_e …end boundaries on y-axis, $Y_e = \{y_{e1}, y_{e2}, \ldots, y_{en}\}$
7. / / x_{max} …maximum coordinate on x-axis
8. / / y_{max} …maximum coordinate on y-axis
9. / / X_{be} …2D Bε-string on x-axis, $X_{be} = \{v_0x_1v_1x_2v_2\ldots v_{n-1}x_nv_n\}$, where v_i = dummy object ε or null string; $i = 0, 1, \ldots, n$; $x_i = c_{bi}$ or c_{ei}, identifies the projection of beging boundary or end boundary of object c_i on x-axis, $i = 1, 2, \ldots, n$
10. // Y_{be} …2D Bε-string on y-axis, $Y_{be} = \{v_0y_1v_1y_2v_2\ldots v_{n-1}y_nv_n\}$, where v_i = dummy object ε or null string; $i = 0, 1, \ldots, n$; $y_i = c_{bi}$ or c_{ei}, identifies the projection of beging boundary or end boundary of object c_i on y-axis, $i = 1, 2, \ldots, n$
11. // S…a sort based on x-axis, $S = \{s_i \mid s_i = x_{bi}c_i$ or $x_{ei}c_i, i = 1, 2, \ldots, 2n\}$
12. // T…a sort based on y-axis, $T = \{t_i \mid t_i = y_{bi}c_i$ or $y_{ei}c_i, i = 1, 2, \ldots, 2n\}$

continued on following page

Table 2. continued

```
13.  // Combine MBR coordinate and object identifier as a key, sort the input data by
     ascending order
14.  for i = 1 to n
15.      s_i ← x_bi c_i
16.      s_i+n ← x_ei c_i
17.      t_i ← y_bi c_i
18.      t_i+n ← y_ei c_i
19.  Sorting S and T by ascending order
20 /     / Construct 2D Bε-string on x-axis
21       X_be ← "                          // Initilized by a null string
22 i     f x_b of s_1 ≠ 0 then       /     / Insert ε at the leftmost?
23           X_be ← ε
24 f     or i = 1 to 2n-1
25           if type of x in s_i is x_b then  / / convert coordinate to
26               X_be ← X_be c_bi            // boundary symbol
27           else                          // type of x in s_i is x_e
28               X_be ← X_be c_ei
29           if x of s_i ≠ x of s_i+1 then
30               X_be ← X_be ε
31 i     f x_e of s_2n ≠ x_max then        // Insert ε at the rightmost?
32           X_be ← X_be ε
33 /     / Construct 2D Bε-string on y-axis
34       Y_be ← "
35 i     f y_b of t_1 ≠ 0 then             // Insert ε at the bottommost?
36           Y_be ← ε
37 f     or i = 1 to 2n-1
38           if type of y in t_i is y_b then  / / convert coordinate to
39               Y_be ← Y_be c_bi            // boundary symbol
40           else                          // type of y in t_i is y_e
41           Y_be ← Y_be c_ei
42           if y of t_i ≠ y of t_i+1 then
43               Y_be ← Y_be ε
44 i     f y_e of t_2n ≠ y_max then        // Insert ε at the topmost?
45           Y_be ← Y_be ε
46 r     eturn X_be, Y_be
```

N: number of objects in a query image;

Q_x: string length along x-axis in a query image;

Q_y: string length along y-axis in a query image;

L_x: length of LCS string with dummy objects along x-axis;

L_y: length of LCS string with dummy objects along y-axis;

M_x: length of L_x string without dummy object;

M_y: length of L_y string without dummy object;

D_x: length of boundary symbols with spatial relationships in database image based on M_x;

D_y: length of boundary symbols with spatial relationships in database image based on M_y;

W_x: weight adjustment along x-axis, $0 \le Wx \le 1$;

W_y: weight adjustment along y-axis, $0 \le Wy \le 1$, and $Wx + Wy = 1$;

S_x: similarity along x-axis, $0 \le Sx \le 1$,

Table 3. Algorithm to determine LCS length from two strings

2D-Bε-LCS-Length (*Q*, *D*)
1. $m \leftarrow length(Q)$
2. $n \leftarrow length(D)$
3. // Q is a 2D Bε-string of query image, $Q = \{q_i \mid q_i \in$ {dummy object ε or the boundary symbols of objects in query image}, $i = 1, 2, \ldots, m\}$.
4. / / D is a 2D Bε-string of database image, $D = \{d_j \mid d_j \in$ {dummy object ε or the boundary symbols of objects in query image}, $j = 1, 2, \ldots, n\}$.
5. // LCS length is inferred in table W, $W = \{w_{i,j} \mid |w_{i,j}|$ is the LCS length of string $q_1, \ldots, q_i$ and $d_1, \ldots, d_j$. If the last symbol of LCS string is dummy object ε then $w_{i,j} < 0$ else $w_{i,j} \geq 0$, where $i = 0, 1, 2, \ldots, m$, $j = 0, 1, 2, \ldots, n\}$.
6. // Initilize the first column of inferring table W with zeros.
7. f or $I \leftarrow 1$ to m do
8. $w_{i,0} \leftarrow 0$
9. // Initilize the first row of inferring table W with zeros.
10. for $j \leftarrow 0$ to n do
11. $w_{0,j} \leftarrow 0$
12. // From row 1 column 1, infer the value of each cell, row by row and column by column, until every cell has been evaluated.
13. for $i = 1$ to m do
14. f or $j = 1$ to n do
15. // Set current cell value to the value of the left or upper cell which has maximum absolute value.
16. i f $|w_{i-1,j}| \geq |w_{i,j-1}|$ then
17. $w_{i,j} \leftarrow w_{i-1,j}$
18. else
19. $w_{i,j} \leftarrow w_{i,j-1}$
20. // Then check whether the value of q_i and d_j are equal and at least one of q_i and the last symbol on the LCS left-up diagonal path is not a dummy object.
21. if $(q_i = d_j)$ and $((q_i \neq \varepsilon)$ or $(w_{i-1,j-1} \geq 0))$ then
22. // If all symbols are hold, then check whether to follow the left-up diagonal path.
23. if $(|w_{i-1,j-1}| + 1) > |w_{i,j}|$ then
24. $w_{i,j} \leftarrow |w_{i-1,j-1}| + 1$
25. if $q_i = \varepsilon$ then
26. $w_{i,j} \leftarrow -w_{i,j}$
27. return W

Table 4. Algorithm to print LCS string of two 2D Be-Strings

Print-2D-Bε-LCS (*Q*, *W*, *i*, *j*)
1. // Q is a 2D Bε-string of query image, $Q = \{q_i \mid q_i \in$ {dummy object ε or the boundary symbols of objects in query image}, $i = 1, 2, \ldots, m\}$.
2. / / W is the LCS-length inferring table of two 2D Bε-strings, and is induced by the algorithm in Table 3.
3. i f $i = 0$ or $j = 0$ then
4. return
5. i f $|w_{i,j}| = |w_{i-1,j}|$ then
6. Print-2D-Be-LCS(Q, W, i-1, j)
7. e lse if $|w_{i,j}| = |w_{i,j-1}|$ then
8. Print-2D-Be-LCS(Q, W, i, j-1)
9. e lse
10. Print-2D-Be-LCS(Q, W, i-1, j-1)
11. print q_i
12. return

$$S_x = \begin{cases} 1-(Q_x + D_x - 2L_x)/(4N+1) & \text{if } M_x > 0, \\ 0 & \text{if } M_x = 0; \end{cases}$$

S_y: similarity along *y*-axis, $0 \le Sy \le 1$,

$$S_y = \begin{cases} 1-(Q_y + D_y - 2L_y)/(4N+1) & \text{if } M_y > 0, \\ 0 & \text{if } M_y = 0; \end{cases}$$

S: similarity of the query image to a database image,

S = WxSx + WySy, $0 \le S \le 1$.

The LCS string lengths L_x, L_y, M_x, and M_y can be obtained by the algorithm in Table 3 applied along the *x* and *y* axes. The values of M_x and M_y can be determined as the values L_x and L_y minus the number of dummy symbol, ε, in the LCS string, respectively. W_x and W_y stress the importance of similarity on the *x*-axis and *y*-axis, respectively.

As stated in the section, "2D Bε-String Model", an image with *n* objects has no more than 4*n*+1 symbols on each dimension of its 2D Bε-string. 2*n* symbols are boundary symbols and the rest are 2*n*+1 symbols; dummy objects are dispersed among the boundary symbols. Now each symbol is pictured as a bucket. Thus, an image with *n* objects has up to 4*n*+1 buckets. If the symbol does not exist, then the associated bucket is considered empty. For example, query image 2 with three objects, *A*, *B*, and *C*, has at most 4*3+1=13 symbols. The boundary string is $\varepsilon B_b \varepsilon C_b \varepsilon B_e A_b \varepsilon C_e \varepsilon A_e$. Bucket 7 (Figure 4) is empty due to the lack of a dummy object, ε, between B_e and A_b. The end of a boundary string also lacks a dummy object because the ending boundary of the object *A* is the same as the right-edge of the image; therefore, bucket 13 is empty. In another example, shown in database image 4, the boundary string $\varepsilon B_b \varepsilon D_b \varepsilon B_e A_b \varepsilon D_e \varepsilon A_e \varepsilon$ has an empty bucket 7 (Figure 4).

Comparing the query image (shown as Figure 4) with the database image (shown as Figure 5), buckets in the query image are categorized into four classes and the number of buckets in each class in Table 5 is summarized.

Class I: without a dummy object between two boundary symbols in both the query image and the database image: for example, bucket 7 is empty shown in Figure 4 and Figure 5.

Class II: without a dummy object between two boundary symbols in the query image, but with a dummy object of them in the database image: for example, the ending of boundary string, bucket 13, in the query image (Figure 4), but there is a dummy object in bucket 13 of the database image (Figure 5).

Class III: symbols in the query image but not in the database image; for example, the buckets 4, 5, 10 and 11 show on Figure 4 and Figure 5.

Class IV: symbols with same spatial relationships in both query image and database image; for example, the buckets 1, 2, 3, 6, 8, 9 and 12 present on Figure 4 and Figure 5.

Figure 4. Symbols in buckets of Query Image 2

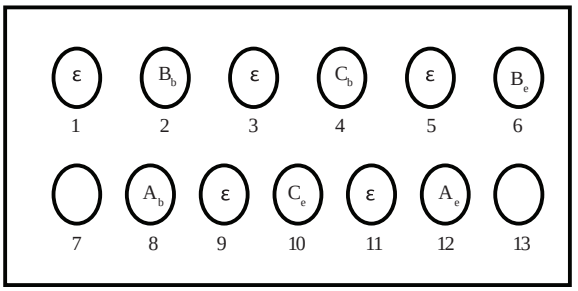


Figure 5. Symbols in bucket of database Image 4

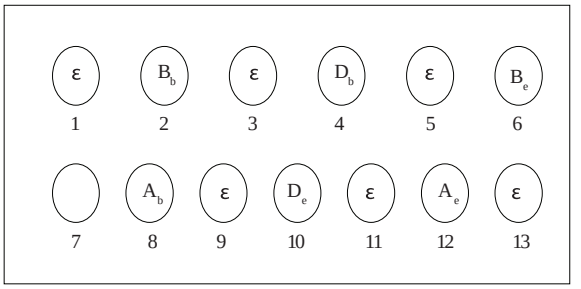


Table 5. Classify by existence of symbol in buckets of a query image

Class	Symbol appears?	The symbol appeared in corresponding bucket in database image?	Number of buckets
I	**No**	**No**	**?**
II N	**o**	**Yes**	**D_x-L_x in *x*-axis, D_y-L_y in *y*-axis**
III	**Yes**	**No**	**Q_x-L_x in *x*-axis, Q_y-L_y in *y*-axis**
IV	**Yes**	**Yes**	**L_x in *x*-axis, L_y in *y*-axis**

Classes I and IV together represent similar parts of a query image and a database image. The evaluated value of similarity can be also obtained from the complementary normalized portion of the total number of buckets in classes II and III. The total number of symbols in class I and IV along the *x*-axis and the *y*-axis are $(4n+1) - (Q_x - L_x + D_x - L_x)$ and $(4n+1) - (Q_y - L_y + D_y - L_y)$, respectively. The similarity on the *x*-axis (S_x) is $1-(Q_x + D_x - 2L_x)/(4n+1)$ and that on the *y*-axis (S_y) is $1-(Q_y + D_y - 2L_y)/(4n+1)$. Clearly, the similarity of the entire image is the sum of S_x and S_y with proper weights adjusting, that is, $S = W_x S_x + W_y S_y$.

Finally, an example is given to detailing the procedure of calculating the similarity. The query image in Figure 3 has three objects, so $N = 3$; the 2D Bε-string is represented as ($\varepsilon A_b \varepsilon B_b \varepsilon A_e C_b \varepsilon C_e \varepsilon B_e \varepsilon$, $\varepsilon B_b \varepsilon A_b \varepsilon B_e C_b \varepsilon C_e \varepsilon A_e \varepsilon$) and the string lengths on the *x*-axis and the *y*-axis, (Q_x, Q_y), are (12, 12), respectively.

The database image in Figure 6 has a 2D Bε-string ($\varepsilon E_b \varepsilon A_b \varepsilon$ $_b \varepsilon A_e \varepsilon C_b \varepsilon F_b \varepsilon E_e \varepsilon C_e \varepsilon B_e \varepsilon F_e \varepsilon$, $\varepsilon E_b \varepsilon B_b \varepsilon E_e \varepsilon A_b \varepsilon B_e C_b \varepsilon F_b \varepsilon C_e \varepsilon A_e \varepsilon F_e \varepsilon$). Table 6 and Table 7 show the LCS-length-inferring table on the *x*-axis and the *y*-axis of the query image and the database image with dummy objects, respectively. The symbols in column q_i and row d_j of Table 6 and Table 7 are the boundary string symbols along the associated axis, with dummy objects in the query image and the database image, respectively. The LCS strings are $\varepsilon A_b \varepsilon B_b \varepsilon A_e C_b \varepsilon C_e \varepsilon B_e \varepsilon$ and $\varepsilon B_b \varepsilon A_b \varepsilon B_e C_b \varepsilon C_e \varepsilon A_e \varepsilon$; thus, the length pair (L_x, L_y) equals (12, 12). The value of (M_x, M_y) can be determined by (L_x, L_y) minus the number

of dummy objects in the LCS string of (L_x, L_y), respectively. The LCS strings of the query image and the database image without dummy objects are $A_bB_bA_eC_bC_eB_e$ and $B_bA_bB_eC_bC_eA_e$; thus, (M_x, M_y) equals (6, 6). The boundary symbols with spatial relationships in the database image, based on strings of M_x and M_y, are $\varepsilon A_b \varepsilon B_b \varepsilon A_e \varepsilon C_b \varepsilon C_e \varepsilon B_e \varepsilon$ and $\varepsilon B_b \varepsilon A_b \varepsilon B_e C_b \varepsilon C_e \varepsilon A_e \varepsilon$; thus, ($D_x$, D_y) equals (13,12). Clearly, the similarity on the *x*-axis and the *y*-axis can be calculated as:

$$\begin{aligned} S_x &= 1-(Q_x + D_x - 2L_x)/(4N+1) \\ &= 1-(12+13-2*12)/(4*3+1) \\ &= 1-(25-24)/13 \\ &= 1-0.0769 \\ &= 0.9231 \end{aligned}$$

$$\begin{aligned} S_y &= 1-(Q_y + D_y - 2L_y)/(4N+1) \\ &= 1-(12+12-2*12)/(4*3+1) \\ &= 1-(24-24)/13 \\ &= 1-0.0 \\ &= 1.0 \end{aligned}$$

The weights adjustments along each axis (Wx, Wy) are given as (0.5, 0.5). Then, for database image 5, the similarity S = 0.5*0.9231+0.5*1.0 = 0.9616.

Rotation and Reflection of an Image

The problems of rotating an image (90, 180, 270 degrees clockwise) or reflecting an image (in the *x*-axis or *y*-axis) are considered in 2-D String, 2D G-string, and 2D C-string. Similarity must be retrieved and evaluated eight times for an image. Thus, a proper transformation must be performed for a string in each dimension. Strings require a sophisticated formula to transform spatial operators, except when reversed (Li & Qu, 1998). Even though 2D B-string does not require cutting objects from an image, it still recalculates their ranks.

If an image is represented in the 2D Bε-string, then the similarity retrieval of rotation and reflection of an image becomes very easy. Before evaluation, the string is only reversed, if necessary. The meaning, a dummy object between two adjoining boundary symbols, is not varied as the image is rotated or reflected because the dummy object is not a spatial operator.

The u and v are assumed to be the boundary strings of an image in the *x*-axis and the *y*-axis, respectively. They have m and n symbols, respectively; that is, $u = \{u_1u_2 \ldots u_{m-1}u_m\}$, $v = \{v_1v_2 \ldots v_{n-1}v_n\}$. The reversed string $u^{-1} = \{u_mu_{m-1} \ldots u_2u_1\}$ is defined for u and $v^{-1} = \{v_nv_{n-1} \ldots v_2v_1\}$ for v (shown as Table 7) summarizes the checklist of 2D Bε-strings between the original image and the image following rotation and/or reflection.

IMPLEMENTATION

A visualized image retrieval system is implemented using the proposed approaches of the 2D Bε-string spatial representation model and the modified LCS similarity evaluation algorithm. The system is divided into four parts. The first part converts MBR coordinate data of the original image into a symbolic picture represented in the 2D Bε-string. Part II randomly generates images, and calls part I to build a symbolic picture database. Each image includes an arbitrary number of objects. Each object is randomly assigned a set of MBR coordinate data. Part III performs similarity retrieval and evaluation. The final part handles the layout of the query image and the user interface interaction. Figure 7 shows the system's architecture.

After the retrieval system is begun (Figure 8), the user can load an image database from storage (Figure 9 and Figure 10), and browse them individually (Figure 11 and Figure 12). The user may place some available icons on the work area of the query image. He then drags them to an appropriate location and scales them appropriately (Figure 13). He hits the 'Search' button after he completes the layout of these icons in the query

Figure 6. Database Image 5

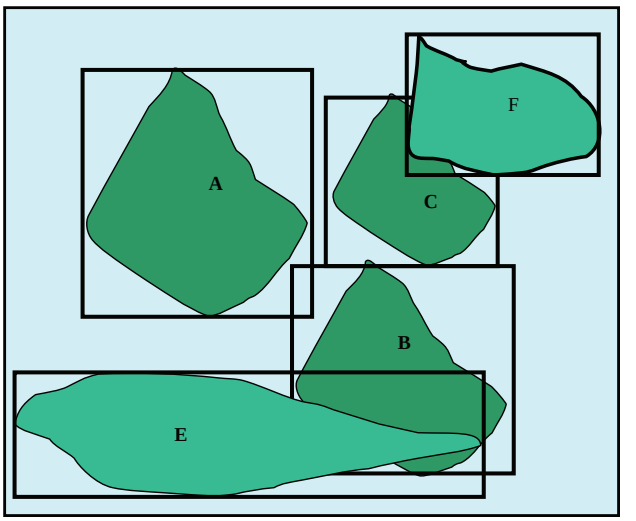


Table 6. LCS length-inferring table with dummy objects on x-axis

W_{ij}		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
		d_i	ε	E_b	ε	A_b	ε	B_b	ε	A_e	ε	C_b	ε	F_b	ε	E_e	ε	C_e	ε	B_e	ε
0	a_i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	ε	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	A_b	0	-1	-1	-1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	ε	0	-1	-1	-1	2	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
4	B_b	0	-1	-1	-1	2	-3	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	ε	0	-1	-1	-1	2	-3	4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
6	A_e	0	-1	-1	-1	2	-3	4	-5	6	6	6	6	6	6	6	6	6	6	6	6
7	C_b	0	-1	-1	-1	2	-3	4	-5	6	6	7	7	7	7	7	7	7	7	7	7
8	ε	0	-1	-1	-1	2	-3	4	-5	6	-7	-7	-8	-8	-8	-8	-8	-8	-8	-8	-8
9	C_e	0	-1	-1	-1	2	-3	4	-5	6	-7	-7	-8	-8	-8	-8	-8	9	9	9	9
10	ε	0	-1	-1	-1	2	-3	4	-5	6	-7	-7	-8	-8	-8	-8	-8	9	-10	-10	-10
11	B_e	0	-1	-1	-1	2	-3	4	-5	6	-7	-7	-8	-8	-8	-8	-8	9	-10	11	11
12	ε	0	-1	-1	-1	2	-3	4	-5	6	-7	-7	-8	-8	-8	-8	-8	9	-10	11	-12

Table 7. LCS length-inferring table with dummy objects on y-axis

$W_{i,j}$		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		d_j	ε	F_b	ε	B_b	ε	F_e	ε	A_b	ε	B_e	C_b	ε	F_b	ε	C_e	ε	A_e	ε	F_e	ε
0	a_i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	ε	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	B_b	0	-1	-1	-1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	ε	0	-1	-1	-1	2	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
4	A_b	0	-1	-1	-1	2	-3	-3	-3	4	4	4	4	4	4	4	4	4	4	4	4	4
5	ε	0	-1	-1	-1	2	-3	-3	-3	4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
6	B_e	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	6	6	6	6	6	6	6	6	6	6
7	C_b	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	7	7	7	7	7	7	7	7	7	7
8	ε	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	7	-8	-8	-8	-8	-8	-8	-8	-8	-8
9	C_e	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	7	-8	-8	-8	9	9	9	9	9	9
10	ε	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	7	-8	-8	-8	9	-10	-10	-10	-10	-10
11	A_e	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	7	-8	-8	-8	9	-10	11	11	11	11
12	ε	0	-1	-1	-1	2	-3	-3	-3	4	-5	6	7	-8	-8	-8	9	-10	11	-12	-12	-12

Table 8. Checklist of 2D Be-Strings after rotation and reflection

	Rotation/reflection	**2D Bε-string**
1	Original image (	u, v)
2	90 degree ckw	(v, u^{-1})
3	180 degree ckw	(u^{-1}, v^{-1})
4	270 degree ckw	(v^{-1}, u)
5	Reflect vs. *x*-axis	(u, v^{-1})
6	Reflect vs. *y*-axis	(u^{-1}, v)
7	90 degree ckw and Reflect vs. *x*-axis	(v, u)
8	90 degree ckw and Reflect vs. *y*-axis	(v^{-1}, u^{-1})

image. The system will transform the query image into a 2D Bε-string and compare it to the database images. The most similar image and its similarity information (Figure 14) are then displayed. Clicking on the '2D Bε-string' tab switches to the 2D Bε-string (Figure 15). The user may also browse less similar images (Figure 16).

FUTURE TRENDS

The 2D Bε-string simplifies the representation of spatial relationships and improves the efficiency of similarity retrieval over those of other 2-D String methodologies. Even so, further research is required.

When an object area in an image differs greatly from its MBR area, the similarity is not

Figure 7. System architecture

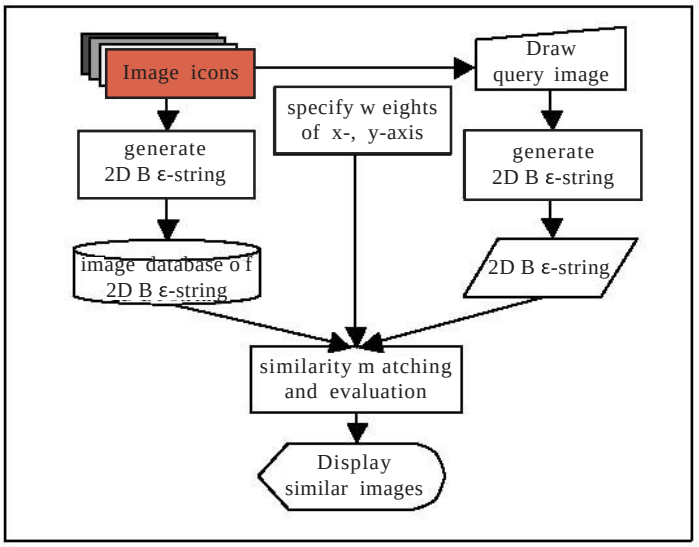
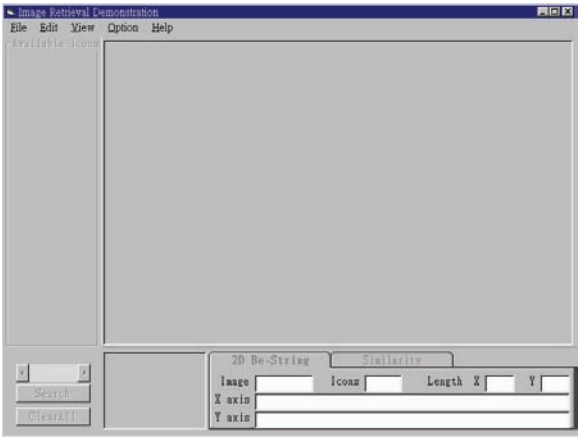


Figure 8. After the image retrieval system has loaded



obvious, whether 2D Bε-string or other 2-D Strings are used. Human vision is such that the size of an object is considered to be one factor that determines similarity. However, 2D Bε-string and other 2-D Strings lose this information while abstracting. The distance between objects is also lost. Consequently, size and distance information in the 2D Bε-string representation model is worthy of further study.

Similarity can be more accurately evaluated than as evaluated here. The original LCS algorithm does not calculate the number of LCS paths, nor does the algorithm presented here. For evaluating similarity of a query image from different images in the database, even for LCS strings with the same length and content, more LCS paths correspond to greater similarity. Thus, modifying the presented similarity retrieval and evaluation approaches to

Figure 9. Loading an image database

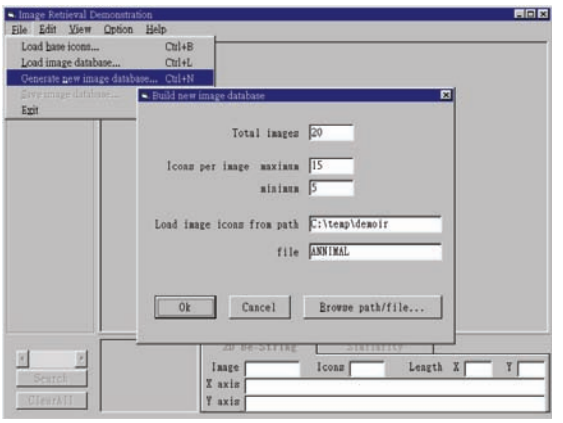


Figure 10. After a database has loaded

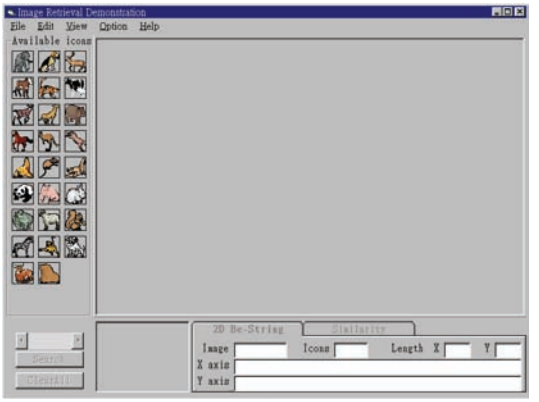


Figure 11. Select to view all database images

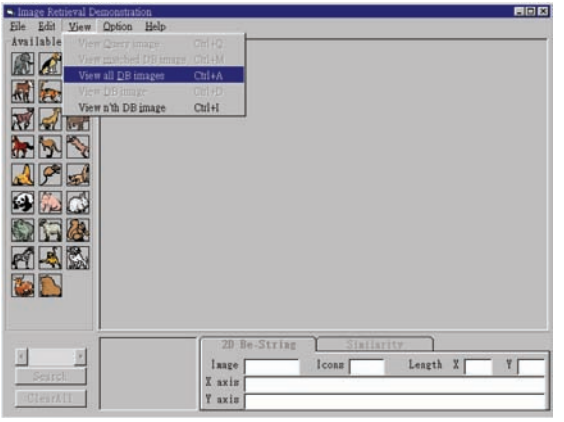


Figure 12. View the first database image

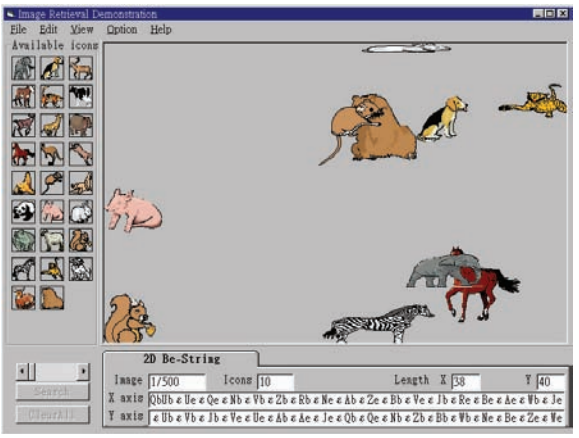


Figure 13. Arrange icons in query image

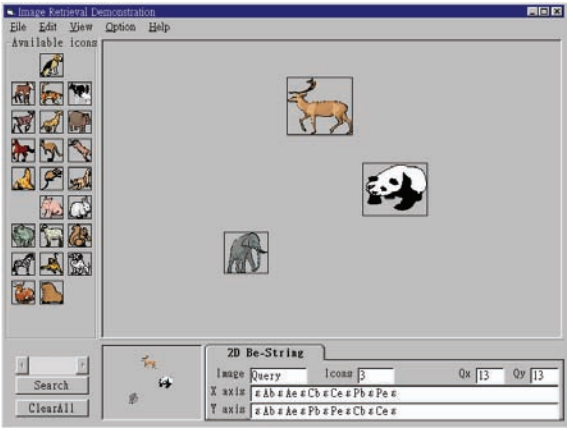


Figure 14. Most similar image and the similarity

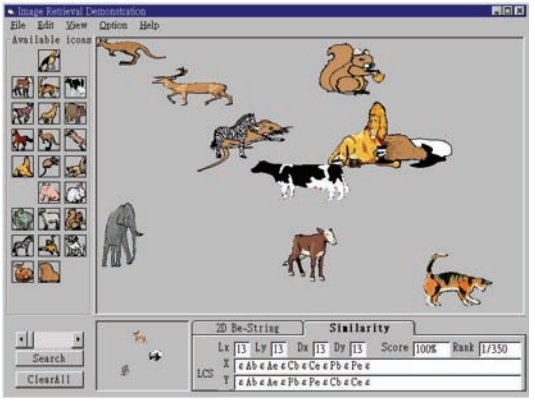


Figure 15. Most similar image and the 2D Be-string

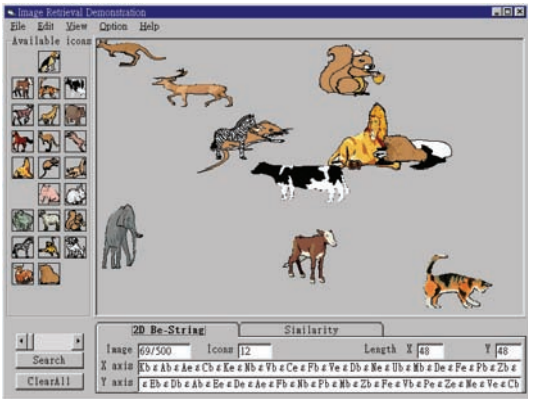
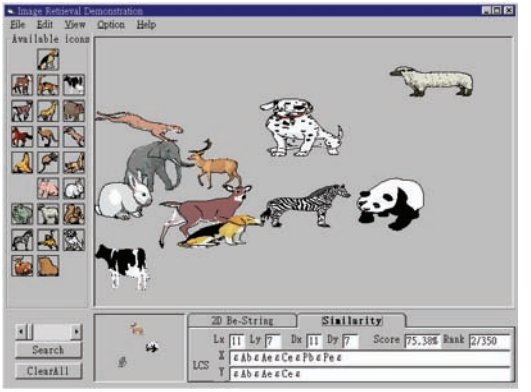


Figure 16. Next similar image and the similarity



account for the number of LCS paths is an area of further study.

The 2D Bε-string representation model and similarity assessment can be easily expanded to retrieve objects in three-dimensions by directly adding the third dimension's string. For video data, adding extra temporal information as the third dimension allows a frame in a video to be indexed. The authors would also like to integrate the movement and velocity of objects in video images into the 2D Bε-string representation model in the near future.

CONCLUSION

This investigation proposes a spatial representation model, called the "2D Bε-string spatial representation" model. The model does not require an icon object to be cut from an image; but it rather represents an object by its MBR boundaries. The model depicts the spatial relationship between two boundary symbols by applying a 'dummy object.' It can, thus, intuitively represent the spatial relationship in an image. Additionally, the 2D Bε-string is directly formed from the object boundaries; thus, an image with n objects needs only between $2n$ and $4n+1$ storage units

in each dimension. The space complexity of 2D Bε-string is $O(n)$.

An algorithm, **Convert-2D-Bε-String**, is introduced to convert an image with MBR coordinate data into a symbolic image, represented in the 2D Bε-string. The space and time complexity of this algorithm is $O(n)$ when it does not consider the sort requirement.

A similarity retrieval algorithm, **2D-Bε-LCS-Length**, is also presented. It is modified from the LCS algorithm for use in the 2D Bε-string representation model. All the space and time complexities of the proposed algorithm are $O(mn)$, where m is the number of icon objects in the query image and n is the number of object icons in the image database.

Moreover, an evaluation process is elucidated, which can evaluate all similarities, regardless of how the LCS string is matched, whether all objects of the query image appear in the database image or whether all of the spatial relationships appear between the pair of images. In retrieving similarities of rotation and reflection, the approach need only to reverse the string and then apply the method of similarity retrieval and evaluation, mentioned above. This process does not require any conversion of spatial operators. It is more efficient and much easier than the previous methodologies.

REFERENCES

Bach, J. R., Gupta, C. F. A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., & Shu, C. (1996). The virage image search engine: An open framework for image management. *SPIE*, 2670.

Beckmann, N., Kriegel, H., Schneider, R., & Seeger, B. (1990). The R*-tree: An efficient and robust access method for points and rectangles. *Proceedings ACM SIGMOD International Conference On the Management of Data*.

Chan, Y. & Chang, C. (n.d.). Image retrieval based on tolerable difference of direction. *Proceedings of the 15th International Conference on Information Networking*, (January 31-February 2). Beppu, Japan.

Chang, S. K., Jungert, E., & Li, Y. (1988). Representation and retrieval of symbolic pictures using generalized 2D string. *Technical Report*, University of Pittsburgh, Pennsylvania, USA.

Chang, S. K., Shi, Q. Y., & Yan, C. W. (1987, May). Iconic indexing by 2-D strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(3), 413-428.

Chien, B. C. (1998). The Reasoning of Rotation and Reflection in Spatial Database. *Systems, Man, and Cybernetics*, IEEE International Conference, (Vol. 2, pp. 1582-1586).

Cormen, T. H., Leiserson, C. E., & Rivest, R.L. (1990). *Introduction to Algorithms.* MIT Press, 314-319.

Guttman, A. (1984). R-tree: A dynamic index structure for spatial searching. *Proceedings of the ACM SIGMOD International Conference on the Management of Data*.

Hsu, F.J., Lee, S. Y., & Lin, B. S. (1999, April). 2D C-tree spatial representation for iconic image. *Journal of Visual Languages & Computing*, 10(2), 147-164.

Huang, P. W. & Jean, Y. R. (1994). Using 2D C^+-string as spatial knowledge representation for image database systems. *Pattern Recognition*, 27(9), 1249-1257.

Kim, B. & Um, K. (1999). 2D+ string: A spatial metadata to reason topological and directional relationships. *Proceedings Of 11th International Conference on Scientific and Statistical Database Management*, (pp. 112-121).

Lee, S. Y. & Hsu, F. J. (1990). 2D C-string: A new spatial knowledge representation for im-

age database systems. *Pattern Recognition*, 23, 1077-1087.

Lee, S. Y., Yang, M. C., & Chen, J. W. (1992). 2D B-string: A spatial knowledge representation for image database systems. *Proceedings ICSC'92 Second International Computer Science Conference*, (pp. 609-615).

Li, X. & Qu, X. (1998). Matching spatial relations using DB-tree for image retrieval. *Pattern Recognition, Proceedings 14th International Conference*, (Vol. 2, pp. 1230-1234).

Liang, E. & Mou, D. (1997). *A Method of Computing Spatial Similarity Between Images.* Graduate Institute of Information Management, Tamkang University, Taiwan.

Liang, E. & Wu, S. (1996). *Similarity Retrieval of Image Database Based On Decomposed Objects.* Graduate Institute of Information Management, Tamkang University, Taiwan.

Liang, E. & You, G. (1999). *Similarity Retrieval Using String Matching in Image Database Systems.* Graduate Institute of Information Management, Tamkang University, Taiwan.

Nibalck, W., Barber, R., Wquitz, W., Flickner, M., Glasman, E., Yanker, D. P. P., Faloutsos, C., & Taubin, G. (1993). The QBIC Project: Querying images by content using color, texture and shape. *SPIE*, 1908.

Orlandic, R. & Yu, B. (2000). A study of MBR-based spatial access methods: How well they perform in high-dimensional spaces. *Proceedings Of International Database Engineering and Application Symposium*, (pp. 306-315).

Papadias, D. & Theodoridis, Y. (1995). Topological relations in the world of minimum bounding rectangles: A study with R-trees. *Proceedings of ACM SIGMOD Conference*, San Jose, CA, USA (pp. 71-79).

Samet, H. (1989). *Applications of Spatial Data Structures, Computer Graphics, Image Processing and GIS*. Addison-Wesley.

Sipser, M. (1997). *Introduction to the Theory of Computation*. PWS Publishing Company, 245-253.

Song, J.W., Whang, K., Lee, Y., & Kim, S.W. (1999). The clustering property of corner transformation for spatial database applications. *Proceedings of the 23rd Annual International Computer Software and Application Conference (COMPSAC'99)*, (pp. 28-35).

Zimbrao, G. & De Souza, J. M. (1998). A raster approximation for the processing of spatial joins. *Proceedings Of the 24th Annual International Conference on VLDB*, New York, USA (pp. 558-569).

This work was previously published in Multimedia Systems and Content-Based Image Retrieval, edited by S. Deb, pp. 277-301, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section 8

Emerging Trends

This section highlights research potential within the field of multimedia technologies while exploring uncharted areas of study for the advancement of the discipline. Introducing this section are selections addressing the need for universal access to multimedia. Additional selections discuss the future of multimedia education, advances in multimedia transmission on the Internet, and the possibilities and limitations of multimedia content protection. These contributions, which conclude this exhaustive, multi-volume set, provide emerging trends and suggestions for future research within this rapidly expanding discipline.

Chapter 8.1
Universal Multimedia Access

Andrea Cavallaro
Queen Mary, University of London, UK

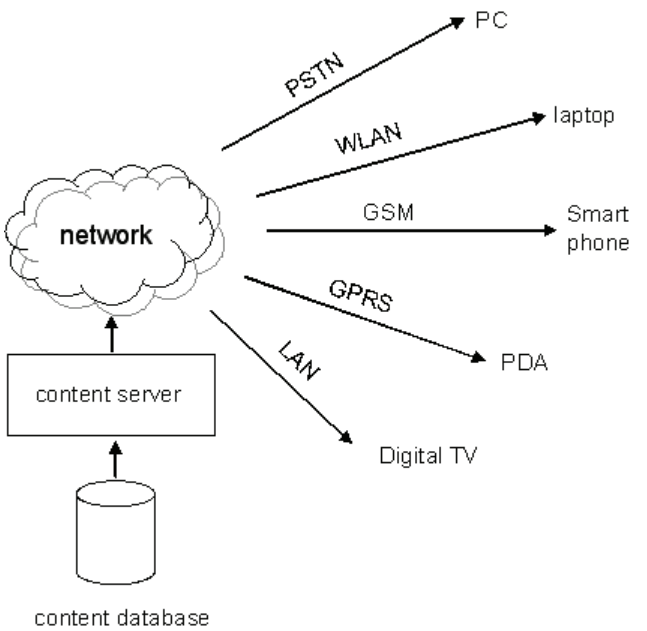
INTRODUCTION

The diffusion of network appliances such as cellular phones, personal digital assistants (PDAs), and handheld computers creates a new challenge for multimedia content delivery: how to adapt the media transmission to various device capabilities, network characteristics, and user preferences. Each device is characterized by certain display capabilities and processing power. Moreover, such appliances are connected through different types of networks with diverse bandwidths. Finally, users with different preferences access the same multimedia content. To cope with the challenge of delivering content to such a variety of conditions while maximizing user satisfaction, multimedia content needs to be adapted to the needs of the specific application, to the capabilities of the connected terminal and network, and to the preferences of the user (Mohan, Smith, & Li, 1999a; Van Beek, Smith, Ebrahimi, Suzuki, & Askelof, 2003). This adaptation enabling seamless access to multimedia content anywhere and anytime is known as universal multimedia access (UMA). The UMA framework is depicted in Figure 1. Three main strategies for adaptive multimedia content delivery have been proposed, namely, the info pyramid, scalable coding, and transcoding. These strategies, emerging trends in UMA and standardization activities, are discussed in the following sections.

INFO PYRAMID

Traditional solutions to multimedia adaptation encode and store multimedia content in a variety of modalities and formats that are expected to fit possible terminals and networks (Li, Mohan, & Smith, 1998). The most adequate version is then selected for delivery according to the network and hardware characteristics of the specific appliance. The advantage of this approach is speed of access because the content is already available and does not need to undergo any transformations. On the other hand, the limitation of this approach is the difficulty of generating a distinct content version for each profile of capabilities for the large variety of terminals and networks currently available. A general framework for managing and manipulat-

Figure 1. Universal multimedia-access framework (PSTN: Public Switched Telephone Network; WLAN: Wireless LAN; LAN: Local Area Network; GSM: Global System for Mobile Communications; GPRS: General Packet Radio Service)

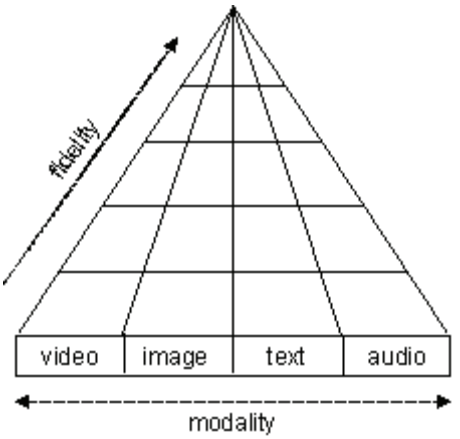


ing media objects is the info pyramid. The info pyramid manages different versions, or variations, of media objects with different modalities (e.g., video, image, text, and audio) and fidelities (summarized, compressed, and scaled variations). Moreover, it defines methods for manipulating, translating, transcoding, and generating the content (Smith, Mohan, & Li, 1999b). When a client device requests a multimedia document, the server selects and delivers the most appropriate variation. The selection is based on network characteristics and terminal capabilities, such as display size, frame rate, color depth, and storage capacity. The info pyramid of a media object is defined as a collection of the different variations of that media object, as shown in Figure 2. A content value score is then associated to each media object. The value score is assigned manually or based on some automatic measure, such as the entropy. Finally, the most appropriate media object is selected by maximizing the total content value for a set of device and/or network constraints. Utility-based frameworks are generally developed for the selection mechanism. In Mohan, Smith, and Li, (1999), the rate-distortion framework is generalized to a value-resource framework by treating different variations of a content item as different compressions, and different client resources as different bit rates. With the info pyramid approach, higher quality or higher resolution bit streams repeat the information already contained in lower quality or resolution streams. Then additional information is added to manage the streams. For these reasons, the info pyramid is not efficient. To overcome this problem, the redundancy should be removed by coding multiple fidelity levels into a single stream, as described in the next section.

SCALABLE CODING

As opposed to the info pyramid, scalable coding processes multimedia content only once. Lower qualities or lower resolutions of the same content

Figure 2. Multimodal representation of a media object as a collection of different variations of the same object in the info-pyramid approach

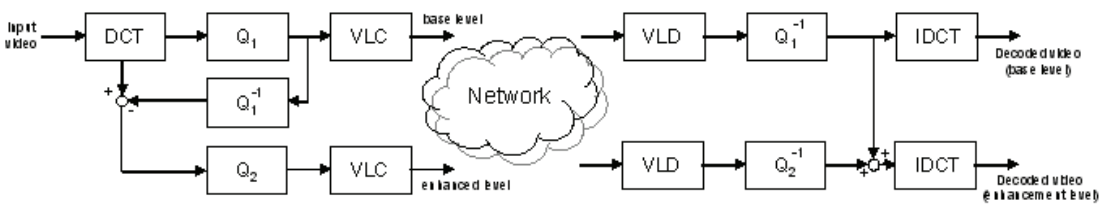


are then obtained by truncating certain layers or bits from the original stream (Wang, Osterman, & Zhang, 2002). In the case of video, basic modes of scalability include quality scalability, spatial scalability, temporal scalability, and frequency scalability. These basic scalability schemes can be combined to reach fine-granularity scalability, such as in MPEG-4 FGS (Fine Granularity Scalability) (Motion Pictures Expert Group; Li, 2001). Quality or SNR (Signal-to-Noise Radio) scalability is defined as the representation of a video sequence with varying accuracies in the color patterns. This is typically obtained by quantizing the color values with increasingly finer quantization step sizes, as shown in Figure 3. Spatial scalability is the representation of the same video in varying spatial resolutions. Corresponding layered bit streams are usually produced by computing a multiresolution decomposition of the original image. Next, the lowest resolution image is coded directly to produce a first layer. For each successive layer, the image from the previous layer is first interpolated to the new resolution, and then the error between the original and the interpolated image is encoded. Temporal scalability is the representation of the same video at varying temporal resolutions or frame rates. The procedure for producing temporally layered bit streams is similar to the procedure used for spatial scalability, but temporal resampling is used instead of spatial resampling. Frequency scalability includes different frequency components in each layer, with the base layer containing low-frequency components and the other layers containing increasingly high-frequency components. Such decomposition can be achieved via frequency transforms like the DCT (Discrete Cosine Transform) or wavelet transforms.

TRANSCODING

Transcoding is the process of converting a compressed multimedia signal into another compressed signal with different properties (Vetro, Christopoulos, & Sun, 2003). Unlike the info pyramid and scalable coding, transcoding can operate according to the current usage environment on the fly without requiring a priori

Figure 3. Flow diagram of an SNR scalable coder based on two levels. Q1, Q2: quantization (Q2 is finer than Q1), VLC: variable-length coder (IDCT: inverse discrete cosin transform)



knowledge of terminal and network capabilities. Early solutions to transcoding determined the output format based on network and appliance constraints only, independent of the semantics in the content. These solutions are referred to as content-blind transcoding techniques. Content-blind transcoding strategies include spatial resolution reduction, temporal resolution reduction, and bit-rate reduction (Vetro et al.). Spatial resolution reduction affects the size of each frame, thus enabling content distribution to devices with limited display capabilities. Spatial resolution reduction can be obtained by first decoding the video stream and then fully reencoding the reconstructed signal at the new resolution. This approach, referred to as cascaded pixel-domain transcoder, has high memory requirements and is computationally expensive. For this reason, spatial resolution reduction is mostly performed in the frequency domain using motion vector mapping and DCT-domain down-conversion techniques (Shanableh & Ghanbari, 2000; Tan, Liang, & Sun, 2004; Vetro, Hata, Kuwahara, Kalva, & Sekiguchi, 2002). Temporal resolution reduction modifies the frame rate to enable content distribution to devices with limited processing power. Frame-rate reduction is acceptable only when motion activity in the video is limited. When motion activity is high, temporal conversion may limit the impression of motion continuity for the user, thus sensibly reducing the perceived quality. Bit-rate reduction aims at meeting an available channel capacity. As for spatial and temporal resolution reduction, significant complexity savings can be achieved by using simplified frequency-domain architectures. For instance, drift-free MPEG-2-video bit-rate reduction can be performed entirely in the frequency domain by implementing the various modes of motion compensation defined by MPEG-2 in the DCT domain (Assuncao & Ghanbari, 1998). The transcoding strategies described thus far, referred to as intramedia transcoding strategies, do not change the media nature of the input signal. On the other hand, intermedia transcoding (or transmoding) is the process of converting the media input into another media format. Examples of intermedia transcoding include speech-to-text (Morgan & Bourlard, 1995) and video-to-text (Jung, Kim, & Jain, 2004) translation.

SEMANTIC ADAPTATION

The success of UMA applications depends on user satisfaction, which in turn depends on the perceived quality of the content delivered. In order to maximize the perceived video quality, an increasing research effort is aimed at improving coders by taking into account human factors (Lu, Lin, Yang, Ong, & Yao, 2004). The various adaptation methods introduced in the previous sections treat the entire scene uniformly, assuming that people may look at every pixel of the video. In reality, humans do not scan a scene in raster fashion. Our visual attention tends to jump from one point to another. These jumps are called saccades. The saccadic patterns depend on the visual scene as well as on the cognitive task to be performed. For this reason, recent adaptation techniques make use of semantics to minimize the degradation of important image regions (Cavallaro, Steiger, & Ebrahimi, 2003). These techniques attempt to emulate the human visual system to prioritize the visual data in order to improve the performance of the coders. To this end, a scene may be decomposed into objects (Cavallaro, Steiger, & Ebrahimi, 2002) as shown in Figure 4. Then, using object-based temporal scalability (OTS), the frame rate of foreground objects is enhanced so that the foreground has a smoother motion than the background. This is usually achieved by encoding the original video sequence at a low frame rate in a base layer. One or more enhancement layers representing only foreground objects are then encoded so as to achieve a higher frame rate than the base layer. Enhancement frames are coded by predicting from the base layer, followed by overlapping the objects

Figure 4. Automatic decomposition of a scene into background and foreground objects. This decomposition enables semantic adaptation with the separate processing of relevant and less relevant visual information.

of the enhancement layer on the combined frame. In semantic transcoding, optimal quantization parameters and frame skip can be determined for each video object individually (Vetro, Sun, & Wang, 2001). The bit-rate budget for each object is allocated by a difficulty hint, a weight indicating the relative encoding complexity of each object. Frame skip is controlled by a shape hint, which measures the difference between two consecutive shapes to determine whether an object can be temporally down-sampled without visible composition problems. Key objects are selected based on motion activity and bit complexity. Motion activity and spatial activity descriptors are used as well to combine the requantization of DCT coefficients with spatial down-sampling and temporal down-sampling for content-based, hybrid video transcoding (Liang & Tan, 2001).

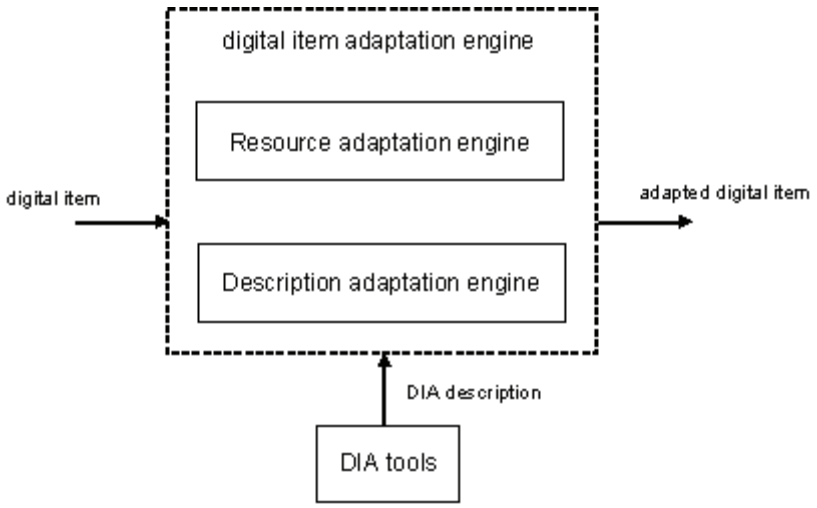
An important open issue in semantic adaptation is quality assessment. Perceptual quality assessment is a difficult task already when dealing with traditional coders such as MPEG-1 and MPEG-2 (Olsson, Stroppiana, & Baina, 1997). When dealing with semantic adaptation and user preferences, the task becomes even more challenging. For this reason, a combination of subjective and objective evaluation techniques is usually employed to compare the performance of different adaptation modalities. Traditional PSNR (Personal Signal-to-Noise Ration) analysis uniformly weighs the contribution of each pixel in an image when computing the mean squared error (MSE). Using this analysis, relevant as well as less relevant parts of an image are given the same importance. To account for the way humans perceive visual information, different parts of an image, or object classes, should be considered (Cucchiara, Grana, & Prati, 2002). Object classes are taken into account through a distortion measure, the semantic mean squared error, which assigns a different weight to each semantic class.

STANDARDS

Enabling access to any multimedia content from any type of terminal or network requires the definition and use of standard tools. In order to achieve interoperable and transparent access to multimedia content, the MPEG standardization committee developed MPEG-21, part 7 (MPEG MDS Group, 2003), which is focused on digital item adaptation (DIA). DIA (Vetro, 2004) aims at providing a set of standardized tools for the adaptation of digital items (Figure 5).

These tools enable the description of the usage environment. This description is based on information about terminal capabilities, network characteristics, user characteristics, and natural

Figure 5. MPEG-21 digital item adaptation architecture



environment characteristics. Terminal capabilities include information on device properties and codec capabilities. Device properties are storage and data I/O (input/output) characteristics, and power-related attributes. Codec capabilities specify the format that a terminal is capable of decoding. Network characteristics include network capabilities and network conditions. Network capabilities define the network's static attributes (e.g., minimum guaranteed bandwidth and maximum capacity), whereas network conditions describe dynamic network parameters, such as delay characteristics, available bandwidth, and error. User characteristics include usage history and user preferences, demonstration preference, accessibility characteristics, and location characteristics. Finally, natural environment characteristics pertain to the physical environmental conditions around the user that affect the way the content is consumed. Noise level and lighting conditions are examples of these characteristics. In addition to the above, natural environment characteristics represent time and location. MPEG-21 DIA specifies only the tools that assist with the adaptation process and not the adaptation engine itself, which is left outside the standard to enable the use of new and improved algorithms.

CONCLUSION

In this article, we discussed the concept of multimedia content delivery anywhere and anytime. In particular, we reviewed different forms for implementing UMA, namely, the info pyramid, scalable coding, and transcoding, and we discussed new adaptation forms based on content semantics. The info pyramid and scalable coding operate when the content is prepared. Content preparation aims at matching possible network and terminal capabilities. For this reason, potential profiles of capabilities need to be known a priori. On the other hand, transcoding takes place at the time of delivery. The input bit stream is converted according to the actual needs of the connected appliance and no prior knowledge is required. However, this flexibility comes at the price of a higher computational load. An emerging trend in UMA is the use of semantics that are being introduced in the adaptation mechanism in order to exploit the characteristics of human perception and maximize user experience.

Because accessing multimedia information anywhere and anytime enables an increase in productivity as well as the improvement of user satisfaction through new multimedia services

and applications, UMA is having a significant impact on large communities, from corporate to private users.

REFERENCES

Assuncao, P. A. A., & Ghanbari, M. (1998). A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit streams. *IEEE Transactions on Circuits and Systems for Video Technology, 8*(8), 953-967.

Cavallaro, A., Steiger, O., & Ebrahimi, T. (2002). Multiple objects tracking in complex scenes. *ACM Multimedia,* 523-532.

Cavallaro, A., Steiger, O., & Ebrahimi, T. (2003). Semantic segmentation and description for video transcoding. *IEEE International Conference on Multimedia and Expo, 3,* 597-600.

Cucchiara, R., Grana, C., & Prati, A. (2002). Semantic transcoding for live video server. *ACM Multimedia,* 223-226.

Jung, K., Kim, K. I., & Jain, A. K. (2004). Text information extraction in images and video: A survey. *Pattern Recognition, 37*(5), 977-997.

Li, C.-S., Mohan, R., & Smith, J. (1998). Multimedia content description in the info pyramid. *IEEE International Conference on Acoustics, Speech and Signal Processing,* 171-178.

Li, W. (2001). Overview of fine granularity scalability in MPEG-4 video standard. *IEEE Transactions on Circuits and Systems for Video Technology, 11*(3), 301-317.

Liang, Y., & Tan, Y.-P. (2001). A new content-based hybrid video transcoding method. *IEEE International Conference on Image Processing, 1,* 429-432.

Lu, Z., Lin, W., Yang, X. K., Ong, E. P., & Yao, S. S. (2004). Spatial selectivity modulated just-noticeable-distortion profile for video. *IEEE International Conference on Acoustic, Speech, and Signal Processing,* 705-708.

Mohan, R., Smith, J., & Li, C.-S. (1999a). Adapting multimedia Internet content for universal access. *IEEE Transactions on Multimedia, 1*(1), 104-114.

Mohan, R., Smith, J., & Li, C.-S. (1999b). Content adaptation framework: Bringing the Internet to information appliances. *IEEE Global Telecommunications Conference,* 2015-2021.

Morgan, N., & Bourlard, H. (1995). Continuous speech recognition. *IEEE Signal Processing Magazine, 12*(3), 24-42.

MPEG MDS Group. (2003). *MPEG-21 multimedia framework, part 7: Digital item adaptation* (ISO/MPEG N5845). Retrieved July 2, 2004, from *http://www.chiariglione.org/mpeg/working_docu ments/mpeg-21/dia/dia_fcd.zip*

Olsson, S., Stroppiana, M., & Baina, J. (1997). Objective methods for assessment of video quality: State of the art. *IEEE Transactions on Broadcasting, 43*(4), 487-495.

Shanableh, T., & Ghanbari, M. (2000). Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats. *IEEE Transactions on Multimedia, 2*(2), 101-110.

Smith, J., Mohan, R., & Li, C.-S. (1999). Scalable multimedia delivery for pervasive computing. *ACM Multimedia, 1,* 131-140.

Tan, Y.-P., Liang, Y., & Sun, H. (2004). On the methods and performances of rational downsizing video transcoding. *Signal Processing: Image Communications, 19,* 47-65.

Van Beek, P., Smith, J., Ebrahimi, T., Suzuki, T., & Askelof, J. (2003). Metadata-driven multimedia access. *IEEE Signal Processing Magazine, 48*(3), 40-52.

Vetro, A. (2004). MPEG-21 digital item adaptation: Enabling universal multimedia access. *IEEE Multimedia,* 84-87.

Vetro, A., Christopoulos, C., & Sun, H. (2003). Video transcoding architectures and techniques: An overview. *IEEE Signal Processing Magazine, 20*(2), 18-29.

Vetro, A., Hata, T., Kuwahara, N., Kalva, N., & Sekiguchi, S.-I. (2002). Complexity-quality analysis of transcoding architectures for reduced spatial resolution. *IEEE Transactions on Consumer Electronics*, 515-521.

Vetro, A., Sun, A., & Wang, Y. (2001). Object-based transcoding for adaptable video content delivery. *IEEE Transactions on Circuits and Systems for Video Technology, 11*(3), 387-401.

Wang, Y., Ostermann, J., & Zhang, Y.-Q. (2002). *Video processing and communications* (1st ed.). Prentice Hall.

KEY TERMS

Info Pyramid: Multimedia data representation based on storing different versions of media objects with different modalities and fidelities.

Intermedia Transcoding: The process of converting the media input into another media format.

Intramedia Transcoding: A transcoding process that does not change the media nature of the input signal.

MPEG: Motion Pictures Expert Group.

MPEG-1: Standard for the coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s. MPEG-1 is the standard on which video CD and MP3 are based.

MPEG-2: Standard for the generic coding of moving pictures and associated audio information. MPEG-2 is the standard on which digital television set-top boxes and DVD are based.

MPEG-4: Standard for multimedia for fixed and mobile Web.

MPEG-7: Multimedia content-description interface. MPEG-7 is the standard for the description and search of audio and visual content.

MPEG-21: Multimedia framework initiative that enables the transparent and augmented use of multimedia resources across a wide range of networks and devices.

Transcoding: The process of converting a compressed multimedia signal into another compressed signal with different properties.

UMA: Universal multimedia access.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 1001-1007, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.2
Towards a Taxonomy of Display Styles for Ubiquitous Multimedia

Florian Ledermann
Vienna University of Technology, Austria

Christian Breiteneder
Vienna University of Technology, Austria

ABSTRACT

In this chapter, a domain independent taxonomy of sign functions rooted in an analysis of physical signs found in public space is presented. This knowledge is necessary for the construction of future multimedia systems that are capable of automatically generating complex yet legible graphical responses from an underlying abstract information space such as a semantic network. The authors take the presence of a sign in the real world as indication for a demand for the information encoded in that sign, and identify the fundamental types of information that are needed to fulfill various tasks. For the information types listed in the taxonomy, strategies for rendering the information to the user in digital mobile multimedia systems are discussed.

INTRODUCTION

Future mobile and ubiquitous multimedia systems will be even more an integrated part of our everyday reality than it is the case today. A digital layer of information will be available in everyday situations and tasks, displayed on mobile devices, blended with existing contents of the real, physical world. Such an "augmented reality" (Azuma et al., 2001) will put into practice recent developments in the area of mobile devices, wireless networking, and ubiquitous information spaces, to be able to provide the right information to the right person at the right time.

The envisioned applications for these kinds of systems are manifold; the scenarios we are thinking of are based on a dense, spatially distributed information space which can be browsed by the user either explicitly (by using

navigation interfaces provided by hardware or software) or implicitly (by moving through space or changing one's intentions, triggering changes in the application's model of the user's *context*). Examples for the information stored in such an information space would be historical anecdotes, routes, and wayfinding information for a tourist guide or road and building information for wayfinding applications. The question of how to encode this information in a suitable and universal way is the subject of ongoing research in the area of semantic modeling (Chen, Perich, Finin, & Joshi, 2004; Reitmayr & Schmalstieg, 2005). For the applications we envision, we will require the information space not only to carry suitable abstract metainformation, but also multimedia content in various forms (images, videos, 3D-models, text, sound) that can be rendered to the user on demand.

Besides solving the remaining technical problems of storage, querying, distribution, and display of that information, which are the subject of some of the other chapters in this book, we have to investigate the consequences of such an omnipresent, ubiquitous computing scenario for the user interfaces of future multimedia applications. Up to now, most research applications have been mainly prototypes targeted towards a specific technical problem or use case; commercial applications mostly focus on and present an interface optimized for a single task (for example, wayfinding). In the mobile and ubiquitous multimedia applications we envision, the user's task and therefore the information that should be displayed cannot be determined in advance, but will be inferred at runtime from various aspects of the user's spatio-temporal context, selecting information and media content from the underlying information space dynamically. To communicate relevant data to the user, determined by her profile, task, and spatio-temporal context, we have to create legible representations of the abstract data retrieved from the information space. A fundamental problem here is that little applicable systematic knowledge exists about the automatic generation of graphical representations of abstract information.

If we want to take the opportunity and clarify rather than obscure by adding another layer of information, the following questions arise: Can we find ways to render the vast amounts of abstract data potentially available in an understandable, meaningful way, without the possibility of designing each possible response or state of such a system individually? Can we replace a part of existing signs in the real world, already leading to "semiotic pollution" (Posner & Schmauks, 1998) in today's cities, with adaptive displays that deliver the information the user needs or might want to have? Can we create systems that will work across a broad range of users, diverse in age, gender, cultural and socio-economical background?

A first step towards versatile systems that can display a broad range of context-sensitive information is to get an overview of which types of information could possibly be communicated. Up to now, researchers focused on single aspects of applications and user interfaces, as for example navigation, but to our knowledge there is no comprehensive overview of what kinds of information can generally occur in mobile information systems. In this article, we present a study that yields such an overview. This overview results in a *taxonomy* that can be used in various ways:

- It can be formalized as a schema for implementing underlying databases or semantic networks
- It can be used by designers to create representative use case scenarios for mobile and ubiquitous multimedia applications
- It can be used by programmers implementing these systems as a list of possible requirements.
- It can be used to systematically search the literature and conduct further research to compile a catalog of display techniques that satisfy the information needs identified.

Such a catalog of techniques, taken from available literature and extended with our own ideas, is presented in the second part of the article

BACKGROUND

Augmented reality blends sensations of the real world with computer-generated output. Already in the early days of this research discipline, its potential to not only add to reality, but also subtract from ("diminished reality") (Mann & Fung, 2002) or change it ("mediated reality") has been recognized. Over the past years, we have created prototypes of mobile augmented reality systems that can be used to roam extensive indoor or outdoor environments. The form factor of these devices has evolved from early back-pack systems (Reitmayr & Schmalstieg, 2004), which prohibited usage over longer time periods or by inexperienced users, to recent PDA-based solutions (Wagner & Schmalstieg, 2003), providing us with a system that can be deployed on a larger scale to untrained and unsupervised users and carried around over an extended time span in an extended environment. Furthermore, on the PDA-class devices, classical and emerging multimedia content formats can be easily integrated, leading to hybrid applications that can make use of different media, matching the needs of the user.

One of our research applications is concerned with outdoor wayfinding in a city (Reitmayr & Schmalstieg, 2004). As can be seen in Figure 1, the augmented reality display provides additional information like directional arrows, a compass, and an indication of the desired target object. After experiments with early ad-hoc prototypes, it became clear that a structured approach to the design of the user interface would be necessary to make our system usable across a wide range of users and tasks. A kind of "toolbox" with different visualization styles is needed to visualize the information in the most suitable way. To design and implement such a toolbox, we need to have an overview of the information needs that might

Figure 1. Our outdoor augmented reality wayfinding system

Directional arrows, landmarks, a compass and location information are superimposed on the view of the real world

occur in our applications, and look for techniques that can successfully fulfill these needs in a flexible, context-dependent way.

Plenty of studies exist that evaluate different display techniques for augmented reality systems. However, we found that the majority of these studies present a novel technique and test the usability of the technique, and do not compare different alternatives for satisfying the same information need. Therefore, these studies were of little direct value for us because they didn't allow us to compare techniques against each other or to find the best technique for a given task. We had to focus on identifying and isolating the proposed techniques, and leave the comparison of techniques against each other for future work. In the future, we will implement some of the proposed techniques and conduct user studies and experiments to be able to compare the techniques to each other.

For conventional 2D diagrams, Chappel and Wilson (1993) present a comparison of different diagram types for various informational purposes. They present a table, listing different tasks (such as, for example, "judging accurate values" or "showing relationships") and for each task, they list the best diagram type according to available cognitive psychology literature. The diagrams discussed include only classical diagram types like pie chart, bar chart or graphs, while we need results in a similar form for recently developed display techniques that can be applied to mobile augmented reality systems.

Some research has been done on the generation of automatic layout for augmented reality displays. Lok and Feiner (2001) present a survey of different automated layout techniques, knowledge that is used by Bell, Feiner, and Höllerer (2001) to present a system for view management for augmented reality. The only information type they are using are labels attached to objects in the view of the user. Nevertheless, their techniques can be applied for controlling the overall layout of an application, once the individual rendering styles for different parts of the display have been chosen.

As the found literature in the field of human-computer-interaction and virtual reality does not answer our questions stated in the introduction, we have to look into other, more theoretical disciplines to find guidelines for the generation of appropriate graphical responses for our systems.

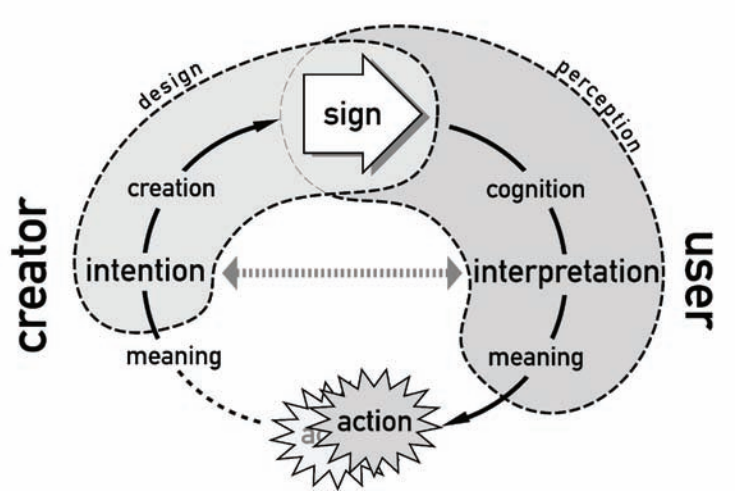
Semiotics and Design

The process that transforms the intention of some agent (software or human) into a legible sign that can be read and understood by users and possibly leads to some action on the user side involves a series of steps: creating a suitable graphical representation for the given intention, placing the created media artifact at a suitable location in the world, identification and perception of the sign by the user, interpreting the sign to extract some meaning and acting according to that meaning. Ideally, the original intention is preserved in this process, and the user acts exactly like the creator intended. However, in the real world these processes are complex, and understanding them is the subject of various scientific disciplines (Figure 2):

- Design theory (Norman, 1990) can teach us how to create aesthetically pleasing and legible signs
- Cognitive psychology (Goldstein, 2004) deals with the perceptual issues involved in sensing and reading
- Semiotics (Eco, 1976) is concerned with the transformation of observed facts into meaning

Generally, the research areas previously mentioned are usually concerned with far less dynamic information than present in the ubiquitous digital applications we are looking for. It is therefore not possible to directly implement the information systems we are envisioning based only on existing knowledge — we first have to examine how these aspects could play together

Figure 2. Sign creation and interpretation



in the context-sensitive applications we want to create. As a first step, we need an overview of what kinds of information can possibly be communicated through signs.

STUDYING REAL-WORLD SIGNS

How can we construct an overview of possible usages of a system we have not yet built? Our hypothesis is that fundamental information needs of our potential users are covered already in the world today, in the form of conventional media and signs. We undertook an exhaustive survey of signs and media artifacts in public space, and from that experience we extracted the core concepts or *atomic functions* of signs in the real world.

Our environments are full of signs—either explicitly and consciously created or left behind without intention. Examples for the first category would be road signs, signposts, labels, and door signs, but also stickers and graffitis, which use public surfaces as ground for articulation and discourse. The signs that are unconsciously created include traces of all kinds, like a path through the grass in a park or the garbage left behind after a barbecue, picnic, or rock concert. Also the design of an object or building can indicate some meaning or suggest some usage that is not explicitly encoded there, but presented as an affordance (Norman, 1990), a feature that is suggesting some way of usage in a more implicit way.

The starting point for our research are signs present in public space. We take existing signs and significant visual features of the environment as indicators for a demand for the information encoded in the sign and/or the individual or political will to create the sign. Therefore, the sign becomes the documentation of the action of its creation, and an indicator of possible actions that can be carried out by using the information that is encoded.

By collecting a large number of examples, we obtained an overview of sign usage in public space and were able to structure intentions and actions into categories, which we could analyze further and relate to each other. In the envisioned ubiquitous augmented reality applications, space and time will be fundamental aspects for structuring the presented information. We therefore focused on signs that are related to spatial or temporal aspects of the world—media created purely for information or the attraction of attention, without any reference to their location or temporal context

Figure 3. Some examples of images taken in our study: (a) annotated safety button; (b) number plate; (c) signposts; (d) roadsign; (e) graffiti; (f) map

(like, for example, advertisements) do not fall in this category.

The collection of examples has been gathered in the city of Vienna, Austria, in public space, public transport facilities, and some public buildings. The research was constrained to include only visual information, and most of the examples were originally photographed with a built-in mobile phone camera. This allowed the spontaneous gathering of new example images in everyday situations, and avoided the necessity to embark for specific "signspotting" trips, which would probably have biased the collection in some direction. Some of the images have been replaced by high-resolution images taken with a consumer digital camera on separate occasions; care has been taken to reproduce the original photo as closely as possible. An unstructured collection of example images is shown in Figure 3.

Obviously, the collection of examples is heavily biased by the photographer's view of the city, his routes, tasks, and knowledge. An improved approach would include several persons with different demographical backgrounds, especially age, cultural and professional background and of varying familiarity with the city. However, our study covers a good part of the explicit signs present in urban space, and allows us to draw conclusions that will be valuable for future research by us and others.

FUNDAMENTAL FUNCTIONS OF SIGNS

In this section, we give an overview of all atomic functions identified in our study. While it is impossible to prove that a given set of categories covers all possible examples without examining every single instance, these categories could already be successfully applied to a number of newly found examples. Therefore, there is some indication that the proposed set of functions covers at least a good part of the use cases that can be found in an urban, public space scenario.

We choose to arrange the functions in five fields, resembling what in our opinion are fundamental aspects of future context sensitive ubiquitous applications: Object metainformation, object relationship information, spatial information, temporal information, and communication. Inside the respective sections, the identified concepts are listed and discussed, together with possible display styles that can be used to render the information in multimedia information systems.

Object Metainformation

Adding metainformation to existing objects in the real world is a fundamental function of both real and digital information systems.

Naming

Naming establishes a linguistic reference for an object in a specific context. The user has to be part of that context to be able to correctly understand the name and identify the referenced object. The context also determines whether the name is unique or not—for example, the name of an institute is unique in the context of a university, but not in a global context. Depending on the user, displayed names have to be chosen appropriately to allow identification.

Identification

Identification is a more technical concept than naming, which allows identifying a specific entity, usually in a global context. Examples would be number plates for cars or street addresses for houses. Note that also in these examples, the identification might need additional parts in a larger context—in a city, the street name is usually unique, but not in a global context, where it has to be prefixed with country and municipality information.

Explanation

Explanation is important if it is not clear from an object's design how to use it, or if the user just wants it for informational purposes. Sometimes it is sufficient to name the object, if the name already implies the mode of operation. A special class of explanation that we identified is *type information*—information about what an object *is*. In contrast to naming, type information denotes the class of an object, and does not provide a reference to a specific instance. (Note that when only a single instance of an object is present in the current context, the type information might also be sufficient to identify the object. Example: "the door" in a room with only a single exit.)

As these three kinds of object-related information mentioned above are mostly textual, the primary problem for displaying it in a digital system is that of automatic layout. The placement, color, and size of labels have to be chosen to be legible, unobtrusive, and not conflicting with other elements of the display. Lok and Feiner (2001) examine different strategies of automatically

Figure 4. Examples for accentuated objects: (a) fire extinguisher; (b) first step of descending stairs; (c) important announcement in public transport system

generating appropriate layouts, knowledge which was used by Bell et al. (2001) to automatically place labels for objects in an augmented reality application.

Accentuation

Accentuation means to emphasize a specific object by increasing its visibility. In the real world, accentuation is mostly performed to permanently improve the visibility of objects or regions for safety reasons by using bright, high contrast colors. In digital systems, image-based methods like partially increasing the contrast or saturation could be used, as well as two- or three-dimensional rendering of overlay graphics. An approach found in some systems (Feiner, Macintyre, & Seligmann, 1993), however never formally evaluated against other techniques, is to superimpose a wireframe model of the object to be highlighted on the object — if the object in question is occluded by other things, dashed lines are used to indicate this. This approach is inspired by technical drawings, where dashed lines are often used to indicate invisible features.

Ownership

While ownership is actually relational information (to be discussed in the next section), linking an owner entity to a specific object, it can often be read as information about the purpose of an object. Examples are the logos of public transport companies on buses. In most cases, the user is not interested in a link to the location of the company, but reads the ownership information as an indication of the object's function.

General Metainformation

Metainformation is often found on device labels to indicate some key properties of the device. Obviously, in digital systems this information can be subject to sophisticated filtering, rendering only the relevant information according to the user's task context. For textual metainformation, the layout considerations discussed above apply.

Status

Display of an object's status is the most dynamic metainformation found in conventional signs—the current state of an object or a subsystem is displayed to the user by using LEDs or alphanumeric displays. In today's cities, this is used for example in public transport systems to display the time until arrival of the next bus.

Status information is, due to its dynamic nature, an example where conventional, physical signs are reaching their limitations. In digital information systems, the possibilities to include dynamic information are much greater. Appropriate filtering has to be applied to prevent information overload and provide only the necessary information to the user. For a discussion of information filtering in an augmented reality context, see Julier, Livingston, Brown, Baillot, and Swan (2000).

Object-Relationship Information

The second type of information we find in various contexts is relating objects to each other. Entities frequently related to each other are people, rooms, buildings, or locations on a map. In most cases, the location of both objects (and the user) determines how the relationship is displayed and what actions can be carried out by the user.

Linking

Linking an object in the real world with another entity is another often-found purpose of signs. In augmented reality applications, one of the two objects (or both) might be virtual objects placed at real world locations. For example, an object in the real world might be linked to a location on a map presented on the user's display.

Rendering a link to the user depends on how the user is supposed to use that information. If the user should be guided from the one object to the other one, arrows can be used to give directional information (see the section on wayfinding below). If the objects are related in some other way, it might be sensible to display the name, an image, or a symbolic representation of the second object, if available, and denote the type of relationship as suitable. If the two objects are close together and both are visible from the users point of view, a straight line can be rendered to connect the objects directly—an approach also used by Bell et al. (2001) to connect labels with the objects they are related to.

Browsing

Browsing means to give the user an overview of all entities that are available for a specific interaction. Real-world examples for browsing opportunities would be signs in the entrance areas of buildings that list all available rooms or persons. The user can choose from that list or look for the name of the entity she is trying to locate.

Computers are frequently used for browsing information. In contrast to the physical world, browsing can be combined with powerful information filtering that passes only relevant information to the user. In most cases, the system will be able to choose the relevant information from the user's context, making browsing only necessary when an explicit choice is to be made by the user.

Spatial Information

The term "navigation" is often used casually for some of the concepts in this section. In our research we found out, however, that we have to break this term down into subconcepts to get an insight into the real motivations and demands of users.

Wayfinding

Wayfinding is what is most often referred to as navigation—finding the way from the current location to a specific target object. Note that for wayfinding only, other aspects of the user's spatial context like overview or orientation can be ignored—the user could be guided by arrows, without having any mental representation of the space she is moving through. In real spaces, wayfinding is supported by arrows and signposts, labeled with the name of the destination object or area. In digital applications, a single, constantly displayed arrow can be used that changes direction as needed.

Overview

Overview supports the ability to build a mental model of the area and is useful for generic wayfinding — finding targets for which no explicit wayfinding information is available, or finding fuzzy targets like areas in a city or district. Also, overview is related to browsing, as it allows looking for new targets and previously unknown locations. Traditionally, overview has been supported by maps (Däßler, 2002). Digital maps offer several new possibilities, like the possibility to mark areas that have been visited by the user before (see the section on trails below).

Orientation

To be useful for wayfinding, overview has to be complemented by orientation, the ability of the user to locate herself on a map or in her mental model of the environment. Maps installed at fixed locations in the world can be augmented with static "You are here" markers, a feature that can be implemented in a dynamic way on a digital map (Vembar, 2004). Overview is also supported by *landmarks*, distinctive visual features of the environment that can be seen from many different

locations in the world. Ruddle (2001) points out the important role of landmarks in virtual environments, which often offer too few distinctive features with the consequence of users feeling lost or disoriented.

Marking Territories

Marking of districts or territories is another example for spatially related information. Real-world examples include road signs or marks on the ground marking the beginning and ending of certain zones (see Figure 5 for example images). One of the problems that conventional signs have is that a human needs to keep track of the current state of the zones she is in as she moves through space.

Spatial Awareness

Ideally, the beginning and ending markings are accompanied by information that provides continuous, ambient feedback of which zone the user is in. This can be found in some buildings, where different areas are marked by using differently colored marks on the walls. Obviously, in digital information systems there are more advanced ways to keep track of and visualize the zones a user is currently in. Continuous feedback, for example in the form of appropriate icons, can be provided to the user on her display, visualizing the currently active zones.

Remote Sensing

A new possibility that emerges with digital multimedia systems is that of remote sensing. By remote sensing, we mean the accessibility of a live video image or audio stream that can be accessed by the user from remote locations. Information provided by remote sensing is less abstract than the other discussed concepts, and opens up the possibility for the user's own interpretation. CCTV cameras installed in public space are an example of remote sensing, although the user group and technical accessibility are limited.

Traces

Traces are often created by crowd behavior and are indicators for usage or demands. Classical examples are paths through the grass in a park, indicating that the provided paths are not sufficient to fulfill the needs of the visitors. In the digital domain, traces can be much more dynamic, collected at each use of the system and annotated with metainformation like date or current task. Some research exists on how traces can be used to aid wayfinding and overview in large virtual environments (Grammenos, Filou, Papadakos, & Stephanidis, 2002; Ruddle, 2005).

Figure 5. Marking of zones: (a) beginning of a speed-limit zone; (b) dashed border surrounding a bus stop; (c) location awareness by colored marking on the wall

Temporal Information

An area where the limitations of conventional signs become clearly visible is information that changes over time. Temporal change has to be marked in advance if the validity of a sign changes over time (for example, parking limitations constrained only to specific times). This additional information can lead to cluttered and overloaded signs (see Figure 3(d)).

Temporal Marking

Temporal marking can be accomplished much easier in digital systems — if the sign is not valid, it can simply be hidden from the users view. Care has to be taken, however, that information that might be relevant for the user in the future (for example, the beginning of a parking limitation) is communicated in advance to allow the user to plan her actions. Which information is relevant to the user in these cases depends highly on the task and activity.

Temporary Change

Similarly, temporary change means the temporary change of a situation (for example, due to construction work) with an undefined ending date. In real world examples, it is usually clearly visible that the change is only temporary and the original state will be restored eventually. If we want to communicate a temporal change in a digital system, this aspect has to be taken into account.

Synchronization

Good examples for synchronization of different parties are traffic lights. Despite their simplicity, traffic lights are among the most complex dynamic information source that can be found in public space. Obviously, the capabilities of future multimedia systems to communicate dynamic information are much greater; therefore, synchronization tasks can probably be adapted dynamically to the current situation.

Sequencing

Synchronization is related to sequencing, where the user is guided through a series of steps to fulfill a task. In real world examples this is usually solved by providing a list of steps that the user is required to take. In digital systems, these steps can be displayed sequentially, advancing to the next step either by explicit user interaction or automatically, if the system can sense the completion of the previous step (for example, by sensing the user's location).

Communication

While signs are always artifacts of communication, signs in the real world are usually only created by legitimate authorities. There are few examples of direct user to user communication—a possibility that can be extended with digital information systems.

Articulation

The surfaces of a city enable articulation in the form of graffiti and posters. While mostly illegal, it is an important property of physical surfaces that they can be altered, extended, or even destroyed. Digital environments are usually much more constrained in what their users are able to do—the rules of the system are often directly mapped to the interaction possibilities that are offered to the user (not taking into account the possibility of hacking the system and bypassing the provided interaction mechanisms). If we replace physical signs by digital content, we should keep in mind that users may want to interact with the information provided, leaving marks, and comments for other users.

Discourse

Discourse through signs and writings involving two or more parties is much rarer observed in public space. The capabilities of networked information systems could improve the ability to support processes of negotiation and communication between multiple parties in public space.

Mapping the Taxonomy

As mentioned above, a linear representation of a taxonomy cannot reproduce a multi-dimensional arrangement of concepts and the relationships between them. To create a more intuitive overview, we have created a 2-dimensional map of the concepts of the taxonomy (Figure 6). Four of the main fields identified above (metainformation, spatial aspects, temporal aspects, communication) are represented in the corners of the map, and the individual concepts are arranged to represent their relation to these fields. In addition, related concepts are linked in the diagram.

RENDERING AND DISPLAY STYLES FOR MOBILE MULTIMEDIA

The following table summarizes the techniques that we identified for the various types of information from the taxonomy. The third column references appropriate literature, where the listed techniques have been discussed or evaluated. The table lists also tasks, for which no appropriate display technique has been presented or evaluated so far. These situations are opportunities for future work, to develop and evaluate techniques that are able to address the communication of the desired information.

CONCLUSION

To support the systematic design of future ubiquitous multimedia applications, we have provided an overview of the types of information that users may demand or content providers may want to communicate. We rooted that overview in a study of sign usage in the real world, taking

Figure 6. An arrangement of the found concepts on a conceptual map

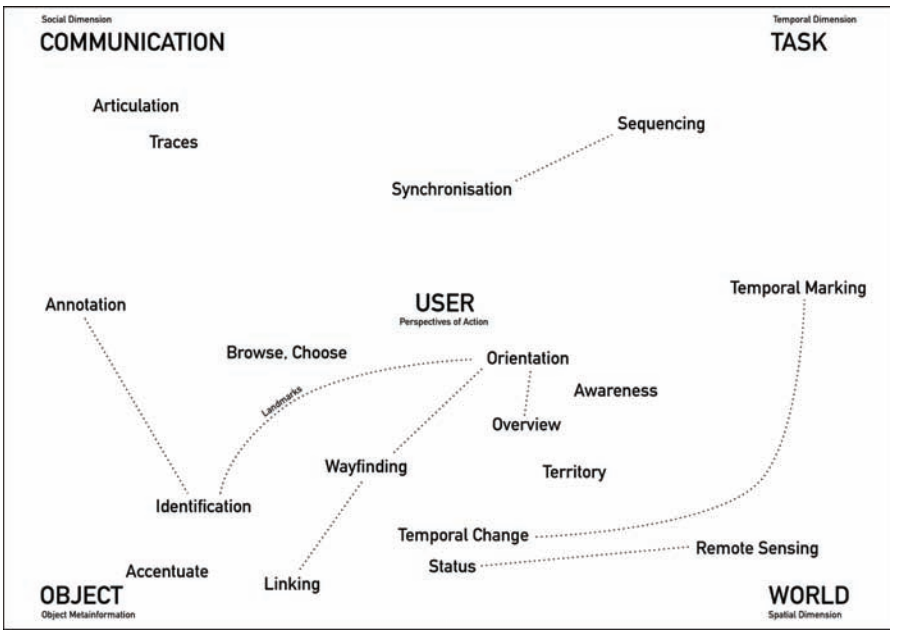


Figure 7.

Task	Technique	References
Labeling: Positioning Labels		Bell et al. (2001)
Metainformation	Information Filtering	Julier et al. (2000)
Highlighting: Visible Objects	Wireframe overlay	Feiner et al. (1993)
Highlighting: Occluded Objects	Cutaway View	Furmanski, Azuma, & Daily et al. (2001)
	Dashed wireframe overlay	Feiner et al. (1993)
Highlighting: Out-of-view Objects		
Linking: Objects to Objects		
Linking: Labels to Objects	Connect with line	Bell et al. (2001)
Linking: Objects to Map		
Navigation: Wayfinding	User–aligned directional arrow	Reitmayr and Schmalstieg (2004)
	Landmarks connected by arrows	Reitmayr and Schmalstieg (2004)
Navigation: Overview	World-in-miniature	Stoakley, Conway, & Pausch et al. (1995)
	Viewer-aligned Map	Diaz and Sims (2003)
Navigation: Orientation	Spatial Audio	Darken and Sibert (1993)
	Landmarks	Darken and Sibert (1993)
	Navigation Grid	Darken and Sibert (1993)
	Breadcrumb Markers	Darken and Sibert (1993)
	Coordinate Feedback	Darken and Sibert (1993)
	Viewer-aligned arrow on map	Vembar (2004)
Territory: Marking		
Traces	Dynamic Trails	Ruddle (2005)
	Breadcrumb Markers	Darken and Sibert (1993)
	Virtual Prints	Grammenos et al. (2002)
Temporal marking		

existing signs as indications for the demand for the information encoded in the sign. From that analysis, we can extrapolate the consequences of bringing that information into the digital domain, which will result in improved possibilities for the display of dynamic information, changing over time and with the context of the user.

While we could identify techniques for rendering some of the information types in digital systems, for some of the identified types of information further research is needed to identify appropriate ways of displaying them to the user. By identifying these "white spots" on our map of display techniques, we provide the basis for future research in the area, targeting exactly those areas where no optimal techniques have been identified so far.

The overview given by the taxonomy may be used by designers of future information systems as a basis for constructing more complex use cases, choosing from the presented scenarios the elements needed for the specific application context. In a (yet to be developed) more formalized way, the presented taxonomy can lay the ground for formal ontologies of tasks and information needs, which could result in more advanced, "semantic" information systems that are able to automatically choose filtering and presentation methods from the user's task and spatio-temporal context.

REFERENCES

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*(6), 34-47.

Bell, B., Feiner, S., & Höllerer, T. (2001). View management for virtual and augmented reality. *Proceedings of the Eurographics Symposion on User Interface Software and Technology 2001 (UIST'01)* (pp. 101-110). New York: ACM Press.

Chappel, H., & Wilson, M. D. (1993). Knowledge-based design of graphical responses. *Proceedings of the ACM International Workshop on Intelligent User Interfaces* (pp. 29-36). New York: ACM Press.

Chen, H., Perich, F., Finin, T., & Joshi, A. (2004). SOUPA: Standard ontology for ubiquitous and pervasive applications. *Proceedings of the International Conference on Mobile and Ubiquitous Systems: Networking and Services*, Boston.

Darken, R. P., & Sibert, J. L. (1993). A toolset for navigation in virtual environments. *Proceedings of the Eurographics Symposion on User Interface Software and Technology 1993 (UIST'93)* (pp. 157-165). New York: ACM Press.

Däßler, R. (2002). Visuelle Kommunikation mit Karten. In A. Engelbert, & M. Herlt (Eds.), *Updates–Visuelle Medienkompetenz*. Würzburg, Germany: Königshauser & Neumann.

Diaz, D. D., & Sims, V. K. (2003). Augmenting virtual environments: The influence of spatial ability on learning from integrated displays. *High Ability Studies, 14*(2), 191-212.

Eco, U. (1976). *Theory of semiotics*. Bloomington: Indiana University Press.

Feiner, S., Macintyre, B., & Seligmann, D. (1993). Knowledge-based augmented reality. *Communications of the ACM, 36*(7), 53-62.

Furmanski, C., Azuma, R., & Daily, M. (2002). Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. *Proceedings of the International Symposion on Mixed and Augmented Reality 2002 (ISMAR'02)* (pp. 215-224). Washington, DC: IEEE Computer Society.

Goldstein, B. E. (2004). *Cognitive psychology* (2nd German ed.). Heidelberg, Germany: Spektrum Akademischer Verlag.

Grammenos, D., Filou, M., Papadakos, P., & Stephanidis, C. (2002). Virtual prints: Leaving trails in virtual environments. *Proceedings of the 8th Eurographics Workshop on Virtual Reality (EGVE'02)* (pp. 131-138). Aire-la-Ville, Switzerland: Eurographics Association.

Julier, S., Livingston, M., Brown, D., Baillot, Y., & Swan, E. (2000). Information filtering for mobile augmented reality. *Proceedings of the International Symposion on Augmented Reality (ISAR) 2000.* Los Alamitos, CA: IEEE Computer Society Press.

Lok, S., & Feiner, S. (2001). A survey of automated layout techniques for information presentations. *Proceedings of SmartGraphics 2001* (pp. 61-68).

Mann, S., & Fung, J. (2002). EyeTap devices for augmented, deliberately diminished, or otherwise altered visual perception of rigid planar patches of real-world scenes. *Presence: Teleoperators and Virtual Environments, 11*(2), 158-175.

Norman, D. (1990). *The design of everyday things.* New York: Doubleday.

Posner, R., & Schmauks, D. (1998). Die Reflektiertheit der Dinge und ihre Darstellung in Bildern. In K. Sachs-Hombach, und K. Rehkämper (Eds.), *Bild–Bildwahrnehmung–Bildverarbeitung. Interdisziplinäre Beiträge zur Bildwissenschaft* (pp. 15-31). Wiesbaden: Deutscher Universitäts-Verlag.

Reitmayr, G., & Schmalstieg, D. (2004). Collaborative augmented reality for outdoor navigation and information browsing. *Proceedings of the Symposium on Location Based Services and TeleCartography.*

Reitmayr, G., & Schmalstieg, D. (2005). Semantic world models for ubiquitous augmented reality. *Proceedings of the Workshop towards Semantic Virtual Environments (SVE'05),* Villars, CH.

Ruddle, R. A. (2001). Navigation: Am I really lost or virtually there? *Engineering Psychology and Cognitive Ergonomics, 6,* 135-142. Burlington, VT: Ashgate.

Ruddle, R. A. (2005). The effect of trails on first-time and subsequent navigation in a virtual environment. *Proceedings of IEEE Virtual Reality 2005 (VR'05)* (pp. 115-122). Bonn, Germany.

Stoakley, R., Conway, M. J., & Pausch, R. (1995). Virtual reality on a WIM: Interactive worlds in miniature. *Conference Proceedings on Human Factors in Computing Systems* (pp. 265-272). Denver, CO: Addison-Wesley.

Vembar, D. (2004). Effect of visual cues on human performance in navigating through a virtual maze. *Proceedings of the Eurographics Symposium on Virtual Environments 2004 (EGVE04).* Aire-la-Ville, Switzerland: Eurographics Association.

Wagner, D., & Schmalstieg, D. (2003). First steps towards handheld augmented reality. *Proceedings of the 7th International Conference on Wearable Computers (ISWC2003),* White Plains, NY.

KEY TERMS

Augmented Reality: Augmented reality (AR) is a field of research in computer science which tries to blend sensations of the real world with computer-generated content. While most AR applications use computer graphics as their primary output, they are not constrained by definition to visual output—audible or tangible representations could also be used. A widely accepted set of requirements of AR applications is given by Azuma (2001):

- AR applications combine sensations of the real world with virtual content.
- AR applications are interactive in real-time
- AR applications are registered in the 3-dimensional space of the real world

Recently, several *mobile* AR systems have been realized as research prototypes, using laptop computers or handheld devices as mobile processing units.

Taxonomy: A taxonomy is a classification of things or concepts, often in a hierarchical manner.

Ubiquitous Computing: The term ubiquitous computing (UbiComp) captures the idea of integrating computers into the environment rather than treating them as distinct objects, which should result in more "natural" forms of interaction with a "smart" environment than current, screen-based user interfaces.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 383-398, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.3
Adaptation and Personalization of Web-Based Multimedia Content

Panagiotis Germanakos
National & Kapodistrian University of Athens, Greece

Constantinos Mourlas
National & Kapodistrian University of Athens, Greece

ABSTRACT

A traditional multimedia system presents the same static content and suggests the same next page to all users, even though they might have widely differing knowledge of the subject. Such a system suffers from an inability to be all things to all people, especially when the user population is relatively diverse. The rapid growth of mobile and wireless communication allowed service providers to develop new ways of interactions, enabling users to become accustomed to new means of multimedia-based service consumption in an anytime, anywhere, and anyhow manner. This chapter investigates the new multi-channel constraints and opportunities emerged by these technologies, as well as the new user-demanding requirements that arise. It further examines the relationship between the adaptation and personalization research considerations, and proposes a three-layer architecture for adaptation and personalization of Web-based multimedia content based on the "new" user profile, with visual, emotional, and cognitive processing parameters incorporated.

INTRODUCTION

Since 1994, the Internet has emerged as a fundamental information and communication medium that has generated extensive enthusiasm. The Internet has been adopted by the mass market more quickly than any other technology over the past century, and is currently providing an electronic connection between progressive entities and millions of users whose age, education, occupation, interest, and income demographics are excellent for sales or multimedia-based service provision.

The explosive growth in the size and use of the World Wide Web, as well as the complicated nature of most Web structures, may lead in orientation difficulties, as users often lose sight of the goal of their inquiry, look for stimulating rather than informative material, or even use the navigational features unwisely. To alleviate such navigational difficulties, researchers have put huge amounts of effort to identify the peculiarities of each user group, and design methodologies and systems that could deliver an adapted and personalized Web-content. To this date, there has not been a concrete definition of personalization. However, the many solutions offering personalization features meet an abstract common goal: to provide users with what they want or need without expecting them to ask for it explicitly (Mulvenna, Anand, & Buchner, 2000). A complete definition of personalization should include parameters and contexts such as user intellectuality, mental capabilities, socio-psychological factors, emotional states, and attention- grabbing strategies, since these could affect the apt collection of users' customization requirements, offering in return the best adaptive environments to the user preferences and demands.

With the emergence of wireless and mobile technologies, new communication platforms and devices, apart from PC-based Internet access, are now emerging, making the delivery of content available through a variety of media. Inevitably, this increases user requirements which are now focused upon an "*anytime, anywhere, and anyhow*" basis. Nowadays, researchers and practitioners not only have to deal with the challenges of adapting to the heterogeneous user needs and user environment issues such as current location and time (Panayiotou & Samaras, 2004), but they also have to face numerous considerations with respect to multi-channel delivery of the applications concerning multimedia, services, entertainment, commerce, and so forth. To this end, personalization techniques exploit Artificial Intelligence, agent-based, and real-time paradigms to give presentation and navigation solutions to the growing user demands and preferences.

This chapter places emphasis on the adaptation of the Web-based multimedia content delivery, starting with an extensive reference to the mobility and wireless emergence that sub-serves the rapid development of the multi-channel multimedia content delivery, and the peculiarities of the user profiling that significantly vary from the desktop to the mobile user. Furthermore, it approaches the existing adaptation (adaptive hypermedia) and personalization (Web personalization) techniques and paradigms that could work together in a coherent and cohesive way, since they are sharing the same goal, to provide the most apt result to the user. Lastly, having analyzed the aforementioned concepts, it defines a three-layer adaptation and personalization Web-based multimedia content architecture that is based on the introduction of a "new" user profile that incorporates user characteristics such as user perceptual preferences, on top of the "traditional" ones, and the semantic multimedia content that includes, amongst others, the perceptual provider characteristics.

MOBILITY EMERGENCE

The rapid development of the wireless and mobile advancements and infrastructures has evidently given "birth" to Mobile Internet. It is considered fundamental to place emphasis on its imperative existence, since statistics show that in the future the related channels will take over as the most sustainable mediums of Web-based (multimedia) content provision. Mobile Internet could be considered as a new kind of front-end access to Web-based content with specific capabilities of delivering on-demand real-time information. Nowadays, many sectors (governmental, private, educational, etc.) start to offer multimedia-based services and information via a variety of service delivery channels apart from the Web (Germanakos, Samaras, & Christodoulou, 2005). Two of

these mobile multimedia-based service delivery channels are mobile telephony and PDAs. These channels become more important considering the much faster growth of the mobile penetration rate compared to desktop-based Internet access. The most significant future development will be the growth of mobile broadband multimedia-based services, once the potential of third generation mobile (3G) and its enhancements, as well other wireless technologies, including W4, RLAN, satellite, and others, is realized. The dissemination of these technologies represents a paradigm shift that enables the emergence of new data multimedia-based services, combining the benefits of broadband with mobility, delivered over high-speed mobile networks and platforms.

Multi-Channel Web-Based Content Delivery Characteristics

"To struggle against the amplification of the digital divide and therefore to think 'user interaction' whatever the age, income, education, experience, and the social condition of the citizen" (Europe's Information Society, 2004).

The specific theme above reveals exactly the need for user-centered multimedia-based service development and personalized content delivery. In many ways, the new technology provides greater opportunities for access. However, there are important problems in determining precisely what users want and need, and how to provide Web-based content in a user-friendly and effective way. User needs are always conditioned by what they already get, or imagine they can get. A channel can change the user perception of a multimedia application; when users have a free choice between different channels to access an application, they will choose the channel that realizes the highest relative value for them. However, separate development of different channels for a single multimedia content (multi-channel delivery) can lead to inconsistencies such as different data formats or interfaces. To overcome the drawbacks of multiple-channel content delivery, the different channels should be integrated and coordinated.

Since successful multimedia-based service delivery depends on a vast range of parameters, there is not a single formula to fit all situations. However, there have been reported particular steps (IDA, 2004) that could guide a provider throughout the channel selection process. Moreover, it should be mentioned that the suitability and usefulness of channels depends on a range of factors, out of which technology is only one element. Additional features that could affect the service channels assessment could be: directness, accessibility and inclusion, speed, security and privacy, and availability. To realize though their potential value, channels need also to be properly implemented and operated.

The design and implementation complexity is rising significantly with the many channels and their varying capabilities and limitations. Network issues include low bandwidth, unreliable connectivity, lack of processing power, limited interface of wireless devices, and user mobility. On the other hand, mobile devices issues include small size, limited processing power, limited memory and storage space, small screens, high latency, and restricted data entry.

Initial Personalization Challenges and Constraints

The needs of mobile users differ significantly from those of desktop users. Getting personalized information "*anytime, anywhere, and anyhow*" is not an easy task. Researchers and practitioners have to take into account new adaptivity axes, along which the personalized design of mobile Web-based content would be built. Such applications should be characterized by flexibility, accessibility, quality, and security in a ubiquitous interoperable manner. User interfaces must be friendlier enabling active involvement (information acquisition), giving the control to the user

(system controllability), providing easy means of navigation and orientation (navigation), tolerating users' errors, supporting system-based and context-oriented correction of users' errors, and finally enabling customization of multi-media and multi-modal user interfaces to particular user needs (De Bra, Aroyo, & Chepegin, 2004; De Bra & Nejdl, 2004). Intelligent techniques have to be implemented that will enable the development of an open Adaptive Mobile Web (De Bra & Nejdl, 2004), having as fundamental characteristics the directness, high connectivity speed, reliability, availability, context-awareness, broadband connection, interoperability, transparency and scalability, expandability, effectiveness, efficiency, personalization, security, and privacy (Lankhorst, Kranenburg, Salden, & Peddemors, 2002; Volokh, 2000).

PERSONALIZATION CONSIDERATIONS IN THE CONTEXT OF DESKTOP AND MOBILE USER

The science behind personalization has undergone tremendous changes in recent years while the basic goal of personalization systems was kept the same, to provide users with what they want or need without requiring them to ask for it explicitly. Personalization is the provision of tailored products, multimedia-based services, Web-based multimedia content, information, or information relating to products or services. Since it is a multi-dimensional and complicated area (covering also recommendation systems, customization, adaptive Web sites, Artificial Intelligence), a universal definition that would cover all its theoretical areas has not been given so far. Nevertheless, most of the definitions that have been given to personalization (Kim, 2002; Wang & Lin, 2002) are converging to the objective that is expressed on the basis of delivering to a group of individuals relevant information that is retrieved, transformed, and/or deduced from information sources in the format and layout as well as specified time intervals.

Comprehensive User Requirements and the Personalization Problem

The user population is not homogeneous, nor should be treated as such. To be able to deliver quality knowledge, systems should be tailored to the needs of individual users providing them personalized and adapted information based on their perceptions, reactions, and demands. Therefore, a serious analysis of user requirements has to be undertaken, documented, and examined, taking into consideration their multi-application to the various delivery channels and devices. Some of the user (customer) requirements and arguments anticipated could be clearly distinguished into Top of the Web (2003) and CAP Gemini Ernst & Young (2004): (a) General User Service Requirements (flexibility: anyhow, anytime, anywhere; accessibility; quality; and security), and (b) Requirements for a Friendly and Effective User Interaction (information acquisition; system controllability; navigation; versatility; errors handling; and personalization).

Although one-to-one multimedia-based service provision may be a functionality of the distant future, user segmentation is a very valuable step in the right direction. User segmentation means that the user population is subdivided, into more or less homogeneous, mutually-exclusive subsets of users who share common user profile characteristics. The subdivisions could be based on: demographic characteristics (i.e. age, gender, urban- or rural-based, region); socio-economic characteristics (i.e. income, class, sector, channel access); psychographic characteristics (i.e. life style, values, sensitivity to new trends); individual physical and psychological characteristics (i.e. disabilities, attitude, loyalty).

The issue of personalization is a complex one with many aspects and viewpoints that need to

be analyzed and resolved. Some of these issues become even more complicated once viewed from a moving user's perspective, in other words when constraints of mobile channels and devices are involved. Such issues include, but are not limited to: what content to present to the user, how to show the content to the user, how to ensure the user's privacy, how to create a global personalization scheme. As clearly viewed, user characteristics and needs, determining user segmentation, and thus provision of the adjustable information delivery, differ according to the circumstances and change over time (Panayiotou and Samaras, 2004).

There are many approaches to address these issues of personalization, but usually, each one is focused upon a specific area, that is, whether this is profile creation, machine learning and pattern matching, data and Web mining, or personalized navigation.

Beyond the "Traditional" User Profiling

One of the key technical issues in developing personalization applications is the problem of how to construct accurate and comprehensive profiles of individual users and how these can be used to identify a user and describe the user behavior, especially if they are moving (Adomavicious & Tuzhilin, 1999). According to Merriam-Webster dictionary, the term profile means "a representation of something in outline". User profile can be thought of as being a set of data representing the significant features of the user. Its objective is the creation of an information base that contains the preferences, characteristics, and activities of the user. A user profile can be built from a set of keywords that describe the user-preferred interest areas compared against information items.

User profiling is becoming more and more important with the introduction of the heterogeneous devices used, especially when published contents provide customized views on information. User profiling can either be static, when it contains information that rarely or never changes (e.g. demographic information), or dynamic, when the data change frequently. Such information is obtained either explicitly, using online registration forms and questionnaires resulting in static user profiles, or implicitly, by recording the navigational behavior and/or the preferences of each user. In the case of implicit acquisition of user data, each user can either be regarded as a member of a group and take up an aggregate user profile or be addressed individually and take up an individual user profile. The data used for constructing a user profile could be distinguished into: (a) the Data Model which could be classified into the demographic model (which describes who the user is), and the transactional model (which describes what the user does); and (b) the Profile Model which could be further classified into the factual profile (containing specific facts about the user derived from transactional data, including the demographic data, such as "the favorite beer of customer X is Beer A"), and the behavioral profile (modeling the behavior of the user using conjunctive rules, such as association or classification rules. The use of rules in profiles provides an intuitive, declarative, and modular way to describe user behavior (Adomavicious & Tuzhilin, 1999)). Additionally, in the case of a mobile user, by user needs is implied both the thematic preferences (i.e., the traditional notion of profile) as well as the characteristics of their personal device called "device profile". Therefore, here, adaptive personalization is concerned with the negotiation of user requirements and device abilities.

But, could the user profiling be considered complete incorporating only these dimensions? Do the designers and developers of multimedia-based services take into consideration the real user preferences in order to provide them a really personalized Web-based multimedia content? Many times this is not the case. How can user profiling be considered complete, and

the preferences derived optimized, if it does not contain parameters related to the user perceptual preference characteristics? We could define User Perceptual Preference Characteristics as all the critical factors that influence the visual, mental, and emotional processes liable of manipulating the newly received information and building upon prior knowledge, that is different for each user or user group. These characteristics determine the visual attention, cognitive, and emotional processing taking place throughout the whole process of accepting an object of perception (stimulus) until the comprehensive response to it. It has to be noted at this point that the user perceptual preference characteristics are directly related to the "traditional" user characteristics since they are affecting the way a user approaches an object of perception.

It is true that nowadays, there are not so many researches that move towards the consideration of user profiling to incorporate optimized parameters taken from the research areas of visual attention processing and cognitive psychology. Some serious attempts have been made on approaching e-learning systems providing adapted content to the students, but most of them are lying to restricted analysis and design methodologies considering particular cognitive learning styles, including Field Independence vs. Field Dependence, Holistic-Analytic, Sensory Preference, Hemispheric Preferences, and Kolb's Learning Style Model (Yuliang & Dean, 1999), applied to identified mental models, such as concept maps, semantic networks, frames, and schemata (Ayersman & Read, 1999). In order to deal with the diversified students' preferences, they are matching the instructional materials and teaching styles with the cognitive styles, and consequently they are satisfying the whole spectrum of the students' cognitive learning styles by offering a personalized Web-based educational content.

A COMPREHENSIVE OVERVIEW OF ADAPTATION AND PERSONALIZATION TECHNIQUES AND PARADIGMS: SIMILARITIES AND DIFFERENCES

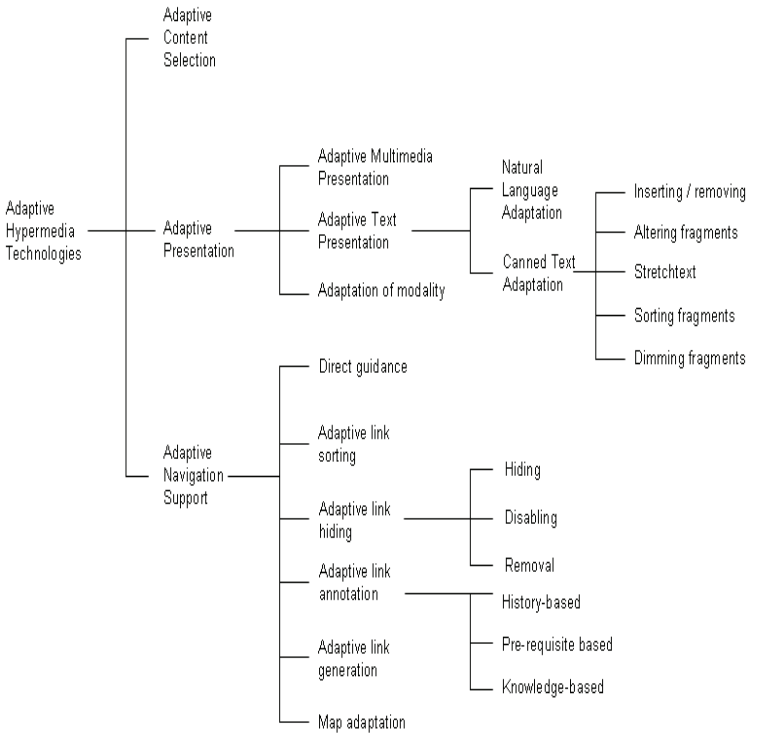
When we are considering adaptation and personalization categories and technologies, we refer to Adaptive Hypermedia and Web Personalization respectively, due to the fact that together these can offer the most optimized adapted content result to the user.

Adaptive Hypermedia Overview

Adaptive Hypermedia is a relatively old and well-established area of research counting three generations: The first "pre-Web" generation of adaptive hypermedia systems explored mainly adaptive presentation and adaptive navigation support and concentrated on modeling user knowledge and goals. The second "Web" generation extended the scope of adaptive hypermedia by exploring adaptive content selection and adaptive recommendation based on modeling user interests. The third "New Adaptive Web" generation moves adaptive hypermedia beyond traditional borders of desktop hypermedia systems embracing such modern Web trends as "mobile Web", "open Web", and "Semantic Web" (Brusilovsky & Maybury, 2002).

Adaptivity is a particular functionality that alleviates navigational difficulties by distinguishing between interactions of different users within the information space (De Bra & Nejdl, 2004; Eklund & Sinclair, 2000). Adaptive Hypermedia Systems employ adaptivity by manipulating the link structure or by altering the presentation of information, based on a basis of a dynamic understanding of the individual user, represented in an explicit user model (Brusilovsky, 1996; De Bra et al., 1999; Eklund, & Sinclair, 2000). In 1996, Brusilovsky identified four user characteristics to which an

Figure 1. Adaptive hypermedia techniques

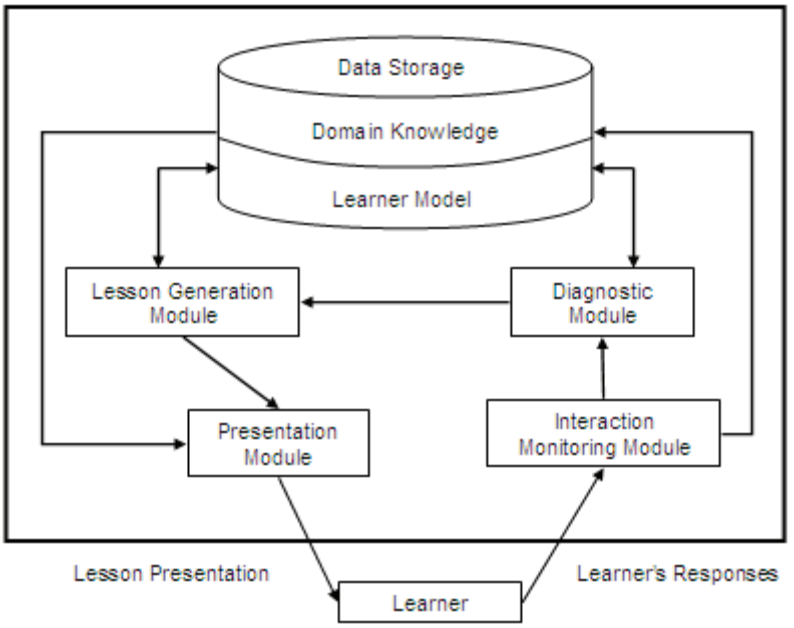


Adaptive Hypermedia System should adapt. These were: user's knowledge, goals, background and hypertext experience, and user's preferences. In 2001, further two sources of adaptation were added to this list, user's interests and individual traits, while a third source of different nature having to deal with the user's environment had also been identified.

Generally, Adaptive Hypermedia Systems can be useful in application areas where the hyperspace is reasonably large and the user population is relatively diverse in terms of the above user characteristics. A review by Brusilovsky has identified six specific application areas for adaptive hypermedia systems since 1996 (Brusilovsky, 2001). These are: educational hypermedia, online information systems, information retrieval systems, institutional hypermedia, and systems for managing personalized view in information spaces. Educational hypermedia and on-line information systems are the most popular, accounting for about two- thirds of the research efforts in adaptive hypermedia. Adaptation effects vary from one system to another. These effects are grouped into three major adaptation technologies: adaptive content selection (De Bra & Nejdl, 2004), adaptive presentation (or content-level adaptation), and adaptive navigation support (or link-level adaptation) (Brusilovsky, 2001; De Bra et al., 1999; Eklund & Sinclair, 2000) and are summarized in Figure 1.

As mentioned earlier, successful adaptation attempts have been made in the e-learning research field to provide the students with adapted content according to their different learning styles or knowledge level and goals. A typical case of such a system could be considered the INSPIRE (Intelligent System for Personalized Instruction in a Remote Environment) architecture, see Figure 2, where throughout its interaction with the

Figure 2. INSPIRE's components and the interaction with the learner



learner, the system dynamically generates lessons that gradually lead to the accomplishment of the learning goals selected by the learner (Papanikolaou, Grigoriadou, Kornilakis, & Magoulas, 2002). INSPSIRE architecture has been designed so as to facilitate knowledge communication between the learner and the system and to support its adaptive functionality.

INSPIRE comprises of five different modules: (a) the *Interaction Monitoring Module* that monitors and handles learner's responses during his/her interaction with the system, (b) the *Learner's Diagnostic Module* that processes data recorded about the learner and decides on how to classify the learner's knowledge, (c) the *Lesson Generation Module* that generates the lesson contents according to learner's knowledge goals and knowledge level, (d) the *Presentation Module* which functions to generate the educational material pages sent to the learner, and (e) the *Data Storage*, which holds the *Domain knowledge* and the *Learner's Model.*

Web Personalization Overview

Web Personalization refers to the whole process of collecting, classifying, and analyzing Web data, and determining based on these the actions that should be performed so that the information is presented in a personalized manner to the user. As inferred from its name, Web Personalization refers to Web applications solely (with popular use in e-business multimedia-based services), and generally is a relatively new area of research. Web personalization is the process of customizing the content and structure of a Web site to the specific needs of each user by taking advantage of the user's navigational behavior. Being a multi-dimensional and complicated area, a universal definition has not been agreed to date. Nevertheless, most of the definitions given to Web personalization (Cingil, Dogac, & Azgin, 2000; Kim, 2002) agree that the steps of the Web personalization process include: (1) the collection of Web data, (2) the modeling and categorization of these data (pre-processing phase), (3) the analysis of the collected data, and

the determination of the actions that should be performed. Moreover, many argue that emotional or mental needs, caused by external influences, should also be taken into account.

Web Personalization could be realized in one of two ways: (a) Web sites that require users to register and provide information about their interests, and (b) Web sites that only require the registration of users so that they can be identified (De Bra et al., 2004). The main motivation points for personalization can be divided into those that are primarily to facilitate the work, and those that are primarily to accommodate social requirements. The former motivational subcategory contains the categories of enabling access to information content, accommodating work goals, and accommodating individual differences, while the latter contains the categories of eliciting an emotional response and expressing identity.

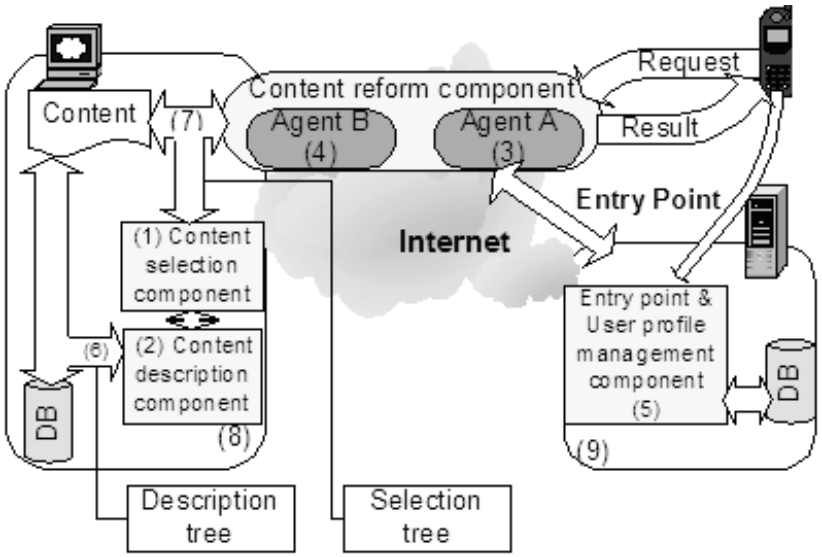
Personalization levels have been classified into: Link Personalization (involves selecting the links that are more relevant to the user, changing the original navigation space by reducing or improving the relationships between nodes), Content Personalization (user interface can present different information for different users providing substantive information in a node, other than link anchors), Context Personalization (the same information (node) can be reached in different situations), Authorized Personalization (different users have different roles and therefore they might have different access authorizations) and Humanized Personalization (involves human computer interaction) (Lankhorst et al., 2002; Rossi, Schwade, & Guimaraes, 2001). The technologies that are employed in order to implement the processing phases mentioned above as well as the Web personalization categories are distinguished into: Content-Based Filtering, Rule-Based Filtering, Collaborative Filtering, Web Usage Mining, Demographic-Based Filtering, Agent Technologies, and Cluster Models (Mobasher, 2002; Pazzani, 1999; Perkowitz & Etzioni, 2003).

The use of the user model is the most evident technical similarity of Adaptive Hypermedia and Web Personalization to achieve their goal. However, the way they maintain the user profile is different; Adaptive Hypermedia requires a continuous interaction with the user, while Web Personalization employs algorithms that continuously follow the users' navigational behavior without any explicit interaction with the user. Technically, two of the adaptation/personalization techniques used are the same. These are adaptive-navigation support (of Adaptive Hypermedia and else referred to as link-level adaptation) and Link Personalization (of Web Personalization) and adaptive presentation (of Adaptive Hypermedia and else referred to as content-level adaptation) and Content Personalization (of Web Personalization).

An example of a Web personalization application for the wireless user is the mPERSONA system, depicted in Figure 3. The mPERSONA system architecture combines existing techniques in a component-based fashion in order to provide a global personalization scheme for the wireless user. The mPERSONA is a flexible and scalable system that focuses towards the new era of wireless Internet and the moving user. The particular architecture uses autonomous and independent components avoiding this way tying up to specific wireless protocols (e.g., WAP). To achieve a high degree of independence and autonomy, mPERSONA is based on mobile agents and mobile computing models such as the "client intercept model" (Panayiotou & Samaras, 2004).

The architectural components are distinguished based on their location and functionality: a) the *Content description* component (Figure 3: 2 & 6), creates and maintains the content's provider metadata structure that describes the actual content, (b) the *Content selection* component (Figure 3: 1 & 7), selects the content that will be presented to the user when "applying" his profile, (c) the *Content reform* component (Figure 3: 3 & 4), reforms and delivers the desired content in the

Figure 3. Detailed view of the mPERSONA architecture



needed (by the user's device) form, and (d) the *User profile management* component (Figure 3: 5), registers and manages user profiles. The user's profile is split into two parts: the device profile (covers the user's devices) and the theme profile (preferences).

A THREE-LAYER ARCHITECTURE FOR ADAPTATION AND PERSONALIZATION OF WEB-BASED MULTIMEDIA CONTENT

Based on the above considerations, a three-layer architecture for adaptation and personalization of Web-based multimedia content will now be presented, trying to convey the essence and the peculiarities encapsulated, and further answering the question why adaptation and personalization of Web-based content is considered vital for the sustainable provision of quality multi-channel Web-based multimedia content/multimedia-based services.

The current architecture, depicted in Figure 4, Adaptation and Personalization of Web-based Multimedia Content Architecture, is composed of three interrelated parts/layers. Each *layer* for the purpose of the infrastructure functionality may be composed of *components*, and each component may be broken down into *elements*, as detailed below:

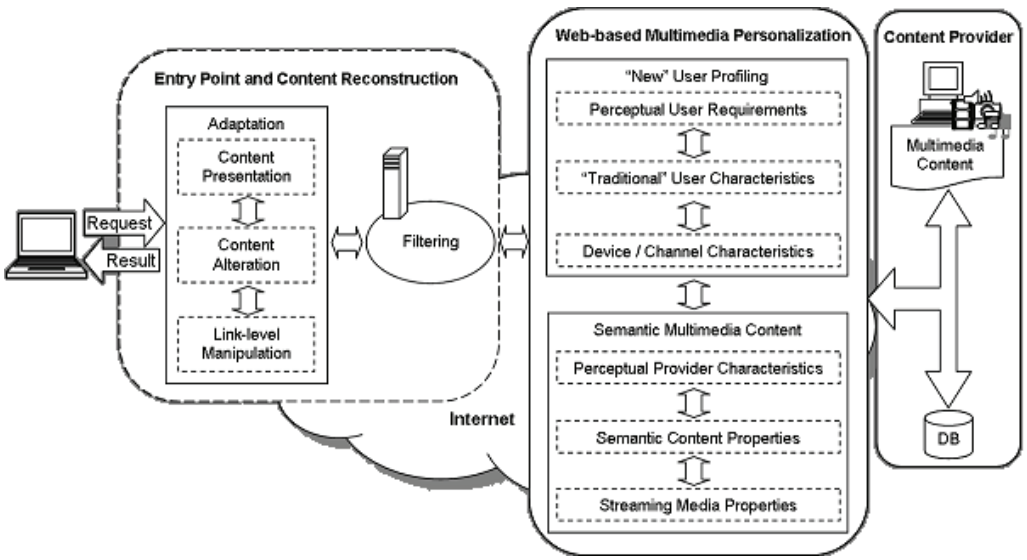
Front-End Layer (Entry Point and Content Reconstruction)

The front-end layer is the primary layer, and it is the user access interface of the system. It is called "Entry Point and Content Reconstruction", and it accepts multi-device requests. It enables the attachment of various devices on the infrastructure (such as mobile phones, PDAs, desktop devices, tablet PC, satellite handset, etc.) identifying their characteristics and preferences as well as the location of the user currently active (personalization/locationbased). It also handles multi-channel requests. Dut to the variety of multi-channel delivery (i.e., over the Web, telephone, interactive kiosks, WAP, MMS, SMS, satellite, and so on),

this layer identifies the different characteristics of the channels. It directly communicates with the middle layer exchanging multi-purpose data. It consists of two components, each one assigned for a different scope:

- **Adaptation:** This component comprises of all the access-control data (for security reasons) and all the information regarding the user profile. These might include user preferences, geographical data, device model, age, business type, native language, context, and so forth. It is the entry point for the user, enabling the login to the architecture. This component is directly communicating with the middle layer where the actual verification and profiling for the user is taking place. Once the whole processing has been completed, it returns the adapted results to the user. It is comprised of three elements:
  - **Content Presentation (or Adaptive Presentation):** It adapts the content of a page to the characteristics of the user according to the user profile and personalization processing. The content is individually generated or assembled from pieces for each user, to contain additional information, pre-requisite information, or comparative explanations by conditionally showing, hiding, highlighting, or dimming fragments on a page. The granularity may vary from word replacement to the substitution of pages to the application of different media.
  - **Content Alteration (or Adaptive Content Selection):** When the user searches for a particular content, that is, related information to his/her profile, the system can adaptively select and prioritize the most relevant items.
  - **Link-Level Manipulation (or Adaptive Navigation Support):** It provides methods that restrict the user's interactions with the content or techniques that aid the user in their understanding of the information space, aiming to provide either orientation or guidance (i.e. adaptive link, adaptive link hiding/an-

Figure 4. Adaptation and personalization of Web-based multimedia content architecture



notation). Orientation informs the user about his/her place in the information space, while guidance is related to a user's goal.

- **Filtering:** This component is considered the main link of the front-layer with the middle layer of the architecture. It actually transmits the data accumulated both directions. It is responsible for making the the low-level reconstruction and filtering of the content, based on the personalization rules created, and to deliver the content for adaptation.

Middle Layer (Web-Based Multimedia Personalization)

The middle layer is the main layer of the architecture and it is called "Web-Based Multimedia Personalization". At this level all the requests are processed. This layer is responsible for the custom tailoring of information to be delivered to the users, taking into consideration their habits and preferences, as well as, for mobile users mostly, their location ("location-based") and time ("time-based") of access. The whole processing varies from security, authentication, user segmentation, multimedia content identification, to provider perceptual characteristics, user perceptions (visual, mental and emotional), and so forth. This layer accepts requests from the front-end and, after the necessary processing, either sends information back or communicates with the next layer (back-end) accordingly. The middle layer is comprised of the following two components:

- **"New" User Profiling:** It contains all the information related to the user, necessary for the Web Personalization processing. It is directly related to the Semantic Multimedia Content component and is composed of three elements:
  - **Perceptual User Requirements:** This is the new element/dimension of the user profile. It contains all the visual attention and cognitive processes (cognitive and emotional processing parameters) that completes the user perception and fulfills the user profile. It is considered a vital element of the user profile since it identifies the aspects of the user that is very difficult to be revealed and measured but, however, might determine his/her exact preferences and lead to a more concrete, accurate, and optimized user segmentation.
  - **"Traditional" User Characteristics:** This element is directly related to the Perceptual User Requirements element and provides the so-called "traditional" characteristics of a user: knowledge, goals, background, experience, preferences, activities, demographic information (age, gender), socio-economic information (income, class, sector, etc.), and so forth. Both elements are completing the user profiling from the user's point of view.
  - **Device/Channel Characteristics:** This element is referring to all the characteristics that referred to the device or channel that the user is using and contains information like: bandwidth, displays, text-writing, connectivity, size, power processing, interface and data entry, memory and storage space, latency (high/low), and battery lifetime. These characteristics are mostly referred to mobile users and are considered important for the formulation of a more integrated user profile, since it determines the technical aspects of it.

- **Semantic Multimedia Content:** This component is based on metadata describing the content (data) available from the Content Provider (back-end layer). In this

way, a common understanding of the data, that is, semantic interoperability and openness is achieved. The data manipulated by the system/architecture is described using metadata that comprises of all needed information to unambiguously describe each piece of data and collections of data. This provides semantic interoperability and a human-friendly description of data. This component is directly related to the "New" User Profile component, providing together the most optimized personalized Web-based multimedia content result. It is consisted of three elements:

- **Perceptual Provider Characteristics:** It identifies the provider characteristics assigned to the Web-based multimedia content or multimedia based service. They are involving all these perceptual elements that the provider has been based upon for the design of the content (i.e., actual content/data of the service, layout preferences, content presentation, etc.)
- **Semantic Content Properties:** This element performs the identification and metadata description of Web-based multimedia content or multi-media-based service based on predetermined ontologies. It is implemented in a transparent manner, removing data duplication and the problem of data consistency.
- **Streaming Media Properties:** It contains data transition mechanisms and the databases. These databases contain the Web-based multimedia content or multimedia-based services as supplied by the provider (without at this point being further manipulated or altered).

Back-End Layer (Content Provider)

This is the last layer of the architecture and is directly connected to the middle layer. It contains transition mechanisms and the databases of Web-based multimedia content or multimedia-based services as supplied by the provider without been through any further manipulation or alteration.

The proposed three-layer architecture for adaptation and personalization of Web-based multimedia content will allow users to receive the Web-based multimedia content or multimedia-based service which they access in an adapted style according to their preferences, increasing in that way efficiency and effectiveness of use.

Implementation Considerations

So far, the functionality and interrelation of three-layer achitecture components that provide adapted and personalized Web-based content have been extensively investigated This section will focus on the concepts and parameters that take part in the construction of a comprehensive user profile, and how these could be used in order to collect all the relevant information. As it has already been mentioned, a lot of research has been done for the implementation of the "traditional" user profiling. Many adaptation and personalization techniques have been developed, and common semantic libraries have been set up that give basically specific and ad-hoc solutions.. However, to our knowledge, implementations that incorporate visual attention, cognitive, and emotional processing parameters to the user profile have not been reported as yet, and such parameters would definitely lead to a comprehensive accumulation of user perceptual preference characteristics and hence, provide users with more sustainable personalized content. Therefore, main emphasis in the following section is given to the construction of the "new" comprehensive user profiling, incorporating these user perceptual preference characteristics mentioned above.

Further examining the middle layer of the proposed architecture and the Perceptual User Requirements element of the "New" User Profiling component, we can see that the User Perceptual Preference Characteristics could be described as a continuous mental processing starting with the perception of an object in the users' attentional visual field (stimulus) and going through a number of cognitive, learning, and emotional processes giving the actual response to that stimulus. This is depicted in Figure 5.

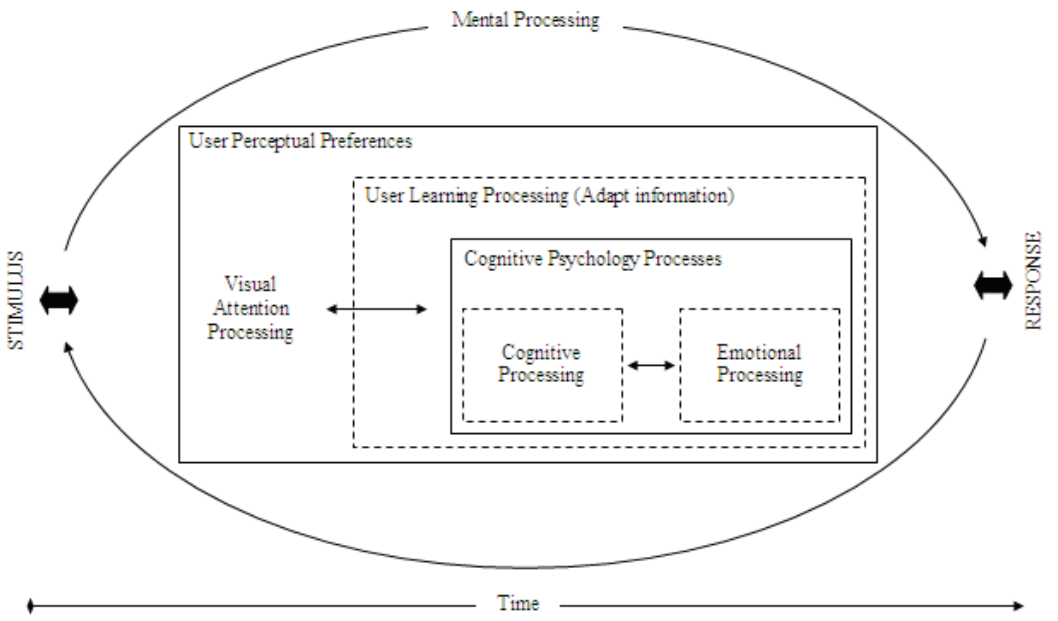
These processes formulate a three-dimensional approach to the problem, as depicted in Figure 6. The three dimensions created are the Learning Styles, the Visual and Cognitive Processing, and the Emotional Processing dimensions.

The *User Learning Processing* dimension is a selection of the most appropriate and technologically feasible learning styles, such as Witkin's Field-Dependent and Field-Independent and Kolb's Learning Styles, being in a position to identify how users transform information into knowledge (constructing new cognitive frames) and if they could be characterized as a converger, diverger, assimilator, accommodator, wholist, analyst, verbalizer, or imager.

The *Visual and Cognitive Processing* dimension is being distinguished from:

- **Visual Attention Processing:** It is composed from the pre-attentive and the limited-capacity stage; the pre-attentive stage of vision subconsciously defines objects from visual primitives, such as lines, curvature, orientation, color, and motion, and allows definition of objects in the visual field. When items pass from the pre-attentive stage to the limited-capacity stage, these items are considered as selected. Interpretation of eye movement data is based on the empirically-validated assumption that when a person is performing a cognitive task, while watching a display, the location of his/her gaze corresponds to the symbol currently being processed in working memory and, moreover, that the eye naturally focuses on areas that are most likely to be informative.

Figure 5. User perceptual preference characteristics

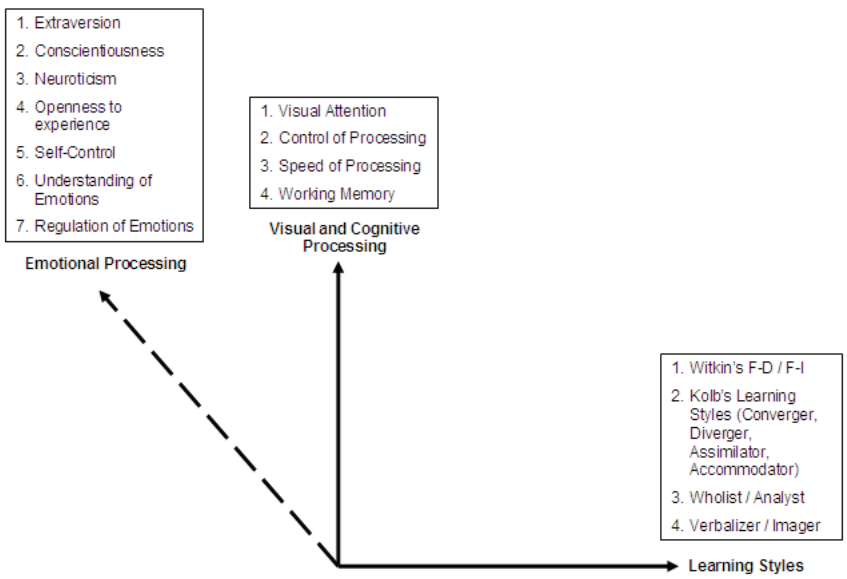


- **Control of Processing:** It refers to the processes that identify and register goal-relevant information and block out dominant or appealing but actually irrelevant information.
- **Speed of Processing:** It refers to the maximum speed at which a given mental act may be efficiently executed (cognitive processing efficiency).
- **Working Memory:** It refers to the processes that enable a person to hold information in an active state while integrating it with other information until the current problem is solved.

The *Emotional Processing* dimension is composed of these parameters that could determine a user's emotional state during the whole response process. This is vital so as to determine the level of adaptation (user needs per time interval) during the interaction process. These parameters include:

- **Extroversion:** Extraverts are sociable, active, self-confident, and uninhibited; while introverts, are withdrawn, shy, and inhibited.
- **Conscientiousness:** Conscientious individuals are organized, ambitious, determined, reliable, and responsible; while individuals low in conscientiousness are distractible, lazy, careless, and impulsive.
- **Neuroticism:** Individuals high in neuroticism are confident, clear-thinking, alert, and content.
- **Open to experience:** Individuals who are open to experience are curious and with wide interests, inventive, original, and artistic; individuals who are not open to experience are conservative, cautious, and mild.
- **Understanding of emotions:** It is the cognitive processing of the emotions; it is the ability of understanding and analysis of the complex emotions and the chain reactions of the emotions, that is, how one emotion generates another.

Figure 6. Three-dimensional approach



- **Regulation of emotions:** It is the control and regulation of personal and other people's emotions for the emotional and intellectual development; it is the human's ability to realize what is hidden behind an emotion, like fear, anxiety, anger, or sadness, and to find each time the most suitable ways to confront them.
- **Self control:** It includes processes referring to the control of attention, the provision of intellectual resources, and the selection of the specialized procedures and skills liable for the evaluation of a problem's results or a decision's uptake; it is a superior control system that coordinates the functioning of other, more specialized control systems.

These parameters must be filtered even more so that the final optimized model is achieved. Once this is established, a series of tests (some in the form of questionnaires and others with real-time interaction metrics) will be constructed which will attempt to reveal users' perceptual preference characteristics. These features, along with the *"Traditional" User Characteristics*, could complete the "New" User Profile, and therefore adaptation and personalization schemes could be adjusted to deliver even more personalized Web-based content accordingly. The next step is to identify what is the correlation between the various users and/or user groups (i.e., to investigate similarities and differences between them) and if it would be feasible to refer to the term "users' segmentation" (i.e., users sharing similar "new" user profiling characteristics). In case the latter is true, personalization mechanisms will be based upon these parameters and considering users' device/channel characteristics, and the semantic content will provide them with the corresponding adapted result. Eventually, this methodology will be implemented with personalization algorithms and paradigms so to automatically gather all the related information and construct the "new" user profiling, giving the users the adapted and personalized result without their actual intervention.

SUMMARY AND FUTURE TRENDS

When referring to adapted multimedia-based services or Web-based multimedia content provision, it is implied that the content adaptation and personalization is based not only on the "traditional" user characteristics, but on a concrete and comprehensive user profiling that covers all the dimensions and parameters of the users preferences. However, knowing the user traditional characteristics and channel/device capabilities, providers can design and offer an apt personalized result. Most of the times, though the providers tend to design multimedia applications based on their own preferences and what they think should be offered. However, the concept of adaptation and personalization is much more complicated than that. This is the reason why until today there is not any sustainable related definition of personalization. A profile can be considered complete when it incorporates the users' perceptual preference characteristics that mostly deal with intrinsic parameters and are very difficult to be technologically measured. Visual attention (that can be thought of as the gateway to conscious perception and memory) and cognitive psychology processes (cognitive and emotional processing parameters) should be in combination investigated and analyzed in a further attempt to complete the desktop and mobile users' preferences.

This chapter made an extensive reference to the mobility emergence and the extensive use of the new channels that tend to satisfy the new user requirements (desktop and mobile) for "*anytime, anyhow, and anywhere*" multimedia-based services and Web-based multimedia content provision in general. The problem of personalization as well as challenges created has been investigated supporting the view of why the provision of adapted content, based on a comprehensive user profile, is considered critical nowadays. Moreover, an Adaptation (Adaptive Hypermedia) and Personalization (Web Personalization) catego-

ries and paradigms review has been presented identifying common grounds and objectives of these two areas. Eventually, a three-layer architecture for the adaptation and personalization of Web-based multimedia content was reviewed, making use of the aforementioned adaptation and personalization concepts and technologies, the new user profiling (that incorporates the user perceptual preference characteristics), as well as the semantic multimedia content..

The basic objective of this chapter was to introduce a combination of concepts coming from different research areas, all of which focus upon the user. It has been attempted to approach the theoretical considerations and technological parameters that can provide the most comprehensive user profiling, supporting the provision of the most apt and optimized adapted and personalized multimedia result.

REFERENCES

Adomavicious, G., & Tuzhilin, A. (1999). User profiling in personalization applications through rule discovery and validation. *Proceedings of the ACM Fifth International Conference on Data Mining and Knowledge Discovery (KDD'99)* (pp. 377-381).

Ayersman, D. J., & Reed, W. M. (1998). Relationships among hypermedia-based mental models and hypermedia knowledge. *Journal of Research on Computing in Education, 30*(3), 222-238.

Brusilovsky, P. (1996). Adaptive hypermedia: An attempt to analyze and generalize. In P. Brusilovsky, P. Kommers, & Streitz (Eds.), *Multimedia, hypermedia, and virtual reality* (pp. 288-304). Berlin: Springer-Verlag.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction, 6*(2-3), 87-129.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction, 11*, 87-110.

Brusilovsky, P., & Maybury, M. T. (2002). From adaptive hypermedia to the adaptive Web. In P. Brusilovsky & M. T. Maybury (Eds.), *Communications of the ACM, 45*(5), *Special Issue on the Adaptive Web*, 31-33.

CAP Gemini Ernst & Young. (2004). Online availability of public services: How is Europe progressing? *European Commission DG Information Society.*

Cingil, I., Dogac, A., & Azgin, A. (2000). A broader approach to personalization. *Communications of the ACM, 43*(8), 136-141.

De Bra, P., Aroyo, L., & Chepegin, V. (2004). The next big thing: Adaptive Web-based systems. *Journal of Digital Information, 5*(1), Article no. 247.

De Bra, P., Brusilovsky, P., & Houben, G. (1999). Adaptive hypermedia: From systems to framework. *ACM Computing Surveys, 31*(4es), 12.

De Bra, P., & Nejdl, W. (2004). Adaptive hypermedia and adaptive Web-based systems. *Proceedings of the Third International Conference (AH 2004),* Springer Lecture Notes in Computer Science, 3137.

Eklund, J., & Sinclair, K. (2000). An empirical appraisal of the effectiveness of adaptive interfaces of instructional systems. *Educational Technology and Society, 3*(4), 165-177.

Europe's Information Society. (2004). *User interaction*. Retrieved from http://europa.eu.int/information_society/activities/egovernment_research/focus/user_interaction/index_en.htm

Germanakos, P., Samaras, G., & Christodoulou, E. (200510-12). Multi-channel delivery of services—the road from e-government to m-government: Further technological challenges and

implications. *Proceedings of the 1st European Conference on Mobile Government (Euro mGov 2005),* Brighton (pp. 210-220).

Interchange of Data between Administrations. (2004). *Multi-channel delivery of e-government services.* Retrieved from http://europa.eu.int/id-abc/

Kim, W. (2002). Personalization: Definition, status, and challenges ahead. *JOT, 1*(1), 29-40.

Lankhorst, M. M., Kranenburg, Salden, A., & Peddemors A. J. H. (2002). Enabling technology for personalizing mobile services. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02): Vol. 3*(3) (p. 87).

Mobasher, B., Dai, H., Luo, T., Nakagawa, M., & Wiltshire, J. (2002). Discovery of aggregate usage profiles for Web personalization. *Data Mining and Knowledge Discovery, 6*(1), 61- 82.

Mulvenna, M. D., Anand, S. S., & Buchner, A. G. (2000). Personalization on the net using Web mining. *Communications of the ACM, 43*(8), 123-125.

Panayiotou, C., & Samaras, G. (2004). mPersona: Personalized portals for the wireless user: An agent approach. *Journal of ACM/ Baltzer Mobile Networking and Applications (MONET), Special Issue on Mobile and Pervasive Commerce, 9*(6), 663-677.

Papanikolaou, K.A., Grigoriadou, M., Kornilakis, H., & Magoulas, G.D. (2002). INSPIRE: An intelligent system for personalized instruction in a remote environment. In S. Reich, M. M. Tzagarakis, & P. M. E. De Bra (Eds.), *OHS/SC/AH 2001, LNCS 2266* (pp. 215-225). Springer-Verlag.

Pazzani, J. M. (1999). A framework for collaborative, content-based, and demographic filtering. *Artificial Intelligence Review, 13*(5-6), 393-408.

Rossi, G., Schwade, D., & Guimaraes, M. R. (2001). Designing personalized Web applications. *ACM Proceedings of the 10th International Conference on World Wide Web* (pp. 275-284).

Top of the Web (2003). Survey on quality and usage of public e-services. Top of the Web. Retrieved from http://www.idt.unisg.ch/org/idt/ceegov.nsf/0/1ae4025175a16a90 c1256df6002a0fef/$FILE/Final_report_2003_quality_and_ usage.pdf

Volokh, E. (2000). Personalization and privacy. *The Communications of the Association for Computing Machinery, 43*(8), 84.

Wang, J., & Lin, J. (2002). Are personalization systems really personal? Effects of conformity in reducing information overload. *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03).* 0-7695-1874-5/03.

Yuliang, L., & Dean, G. (1999). Cognitive styles and distance education. *Online Journal of Distance Learning Administration, 2*(3), Article 005.

This work was previously published in Digital Multimedia Perception and Design, edited by G. Ghinea and S. Y. Chen, pp. 284-304, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.4
New Internet Protocols for Multimedia Transmission

Michael Welzl
University of Innsbruck, Austria

ABSTRACT

This chapter will introduce three new IETF transport layer protocols in support of multimedia data transmission and discuss their usage. First, the stream control transmission protocol (SCTP) will be described; this protocol was originally designed for telephony signaling across the Internet, but it is in fact broadly applicable. Second, UDP-Lite (an even simpler UDP) will be explained; this is an example of a small protocol change that opened a large can of worms. The chapter concludes with an overview of the datagram congestion control protocol (DCCP), a newly devised IETF protocol for the transmission of unreliable (typically real-time multimedia) data streams.

INTRODUCTION

For decades, two transport layer protocols of the TCP/IP suite were almost exclusively used: TCP and UDP. The services that these protocols provide are entirely different, and easy to grasp: while the latter simply makes the "best effort" service of the Internet accessible to applications, TCP reliably transfers a stream of bytes across the network. UDP only has port numbers that make it possible to distinguish between several communicating entities which share the same IP address and a checksum that ensures data integrity, but TCP encompasses a large number of additional functions:

- **Stream-based in-order delivery:** Packets are ordered according to sequence numbers, and only consecutive bytes are delivered
- **Reliability:** Missing packets are detected and retransmitted
- **Flow control:** The receiver is protected against overload with a sliding window scheme
- **Congestion control:** The network is protected against overload by appropriately limiting the window of the sender
- **Connection handling:** Since TCP is a connection oriented protocol, it must have the ability to explicitly set up and tear down connections

- **Full-duplex communication:** An acknowledgment (ACK) can also carry user data; this is usually referred to as "piggybacking"

The importance of these mechanisms varies. A protocol could, for instance, easily do without the full-duplex communication capability; on the other hand, some form of end-to-end congestion control has been identified as an indispensable element of any protocol that is to be used on the Internet (Floyd & Fall, 1999). This does however not mean that there is only one way to carry out congestion control: TCP uses an "additive increase, multiplicative decrease" strategy which essentially probes for the available bandwidth by linearly increasing the rate until a limit is hit (causing a packet to be dropped or a congestion signaling bit to be set), whereupon the rate is reduced by half. There are proposals for congestion control that is fair towards TCP ("TCP-friendly") yet more suitable for multimedia applications because the rate fluctuations are less severe. One notable example is "TCP-friendly rate control (TFRC)" (Floyd, Handley, Padhye, & Widmer, 2000).

TCP does not provide the flexibility that today's applications need: it is neither possible to disable any of its aforementioned functions (in particular reliability, which adds delay but is typically not needed by real-time multimedia applications), nor can a user change the way they work (e.g., influence how congestion control is carried out). UDP, on the other hand, allows for more flexibility, but its feature set is so small that any additional protocol function must be implemented directly within the application that uses it. Sometimes, this is unacceptable — realizing TCP-friendly congestion control, for instance, is difficult, and may not be worth the effort from the perspective of a single application designer. Indeed, even the popular streaming media applications "RealPlayer" and "Windows Media Player" do not appear to properly adapt their rate in response to congestion (Hessler & Welzl, 2005).

In this chapter, we will take a look at three novel IETF protocols that change this situation somewhat: the "stream control transmission protocol (SCTP)," "UDP-Lite," and the "datagram congestion control protocol (DCCP)." While SCTP could also be regarded as some sort of a "TCP++," these three protocols share one notable property: they can emulate the behavior of TCP (or UDP, in the case of UDP-Lite), but with *less* features. The ability to effectively *disable* TCP features is therefore a feature in itself; this gives new meaning to the saying "*less is more*." Historically, SCTP is by far the oldest of these protocols; its main specification (Stewart et al., 2000) was published in 2000, and it is now going through the difficult post-standardization phase of achieving large-scale Internet deployment. Notably, the IETF recommends this protocol for authentication, authorization, and accounting (AAA) in any future IP service networks, and SCTP has been required by the 3rd Generation Partnership Project (3GPP) (Stewart & Xie, 2002, p. 17). UDP-Lite was recently published as a "Proposed Standard" — the same status as SCTP — by the IETF (Larzon, Degermark, Pink, Jonsson, & Fairhurst, 2004), and DCCP has not even reached this status yet; at the time of writing, its specification (Kohler, Handley, & Floyd, 2005) was still an Internet-draft, which is a preliminary type of IETF document. The protocol can be expected to become a Proposed Standard RFC in the near future, and its impact could then become quite significant.

THE STREAM CONTROL TRANSMISSION PROTOCOL (SCTP)

SCTP is the result of an effort to develop an efficient Internet transport protocol for telephony signaling. As such, its features are not directly related to the transmission of multimedia data; it was however understood that it is a protocol of broad use, and SCTP can certainly be advanta-

geous for mobile multimedia if the data are suitable and the protocol is used in an intelligent manner. This is because delay is always an important issue for real-time multimedia applications, and reduced delay is exactly what SCTP can give you. In what follows, we will take a closer look at its main features.

Reliable Out-Of-Order but Potentially Faster Data Delivery

TCP suffers from a problem that is called "head-of-line blocking delay": when packets 0, 2, 3, 4, and 5 reach a TCP receiver, the data contained in packets 2 to 5 will not be delivered to the application until packet 1 arrives. This effect is caused by the requirement to deliver data in order. By allowing applications to relax this constraint (this is reasonable for telephony signaling), SCTP can deliver data faster while providing the reliability that UDP lacks.

Preservation of Message Boundaries

Faster delivery of out-of-order packets is only possible if the data blocks can be clearly identified by the protocol. In other words, embedding such a function in a TCP receiver would not be possible because of its byte stream-oriented nature. Moreover, giving the application the power to control the elementary data units that are transferred ("application layer framing (ALF)") can yield more efficient programs (Clark & Tennenhouse, 1990). This is shown in Figure 1. Here, four application chunks are transmitted in four packets. Without ALF, it is possible that just a couple of bytes from chunk 2 end up in packet 1; if packet 2 (which contains the rest of chunk 2) is lost, however, these bytes are of no use at the receiver until the retransmitted packet 2 arrives. Similarly, the loss of packet 2 can affect chunk 3, rendering the correctly received packet 3 useless until the retransmitted packet 2 arrives.

Efficiently choosing the size of packets as a function of the application chunk size does of course not mean that packets have to be exactly as large as chunks — the same advantage can be gained if the packet size is an integral multiple of the chunk size or vice versa.

Support for Multiple Separate Data Streams

Sometimes, an application may have to transfer more than one logical data stream. Mapping multiple data streams onto a single TCP connection requires some effort from an application and can be inefficient. Figure 2 shows an example scenario where packets are reordered inside the network (this is indicated via the bold numbers underneath the TCP sender and receiver buffers, which represent TCP sequence numbers). Clearly, even when the streams themselves call for in-order data delivery, this is not necessarily the case for segments that belong to different streams, and head-of-line blocking delay can occur — in the figure, chunk 2 from application stream 1 can only be delivered when chunk 1 from the otherwise unrelated application stream 2 arrives. This problem is eliminated by the multiple stream support feature of SCTP.

Another common solution to this problem is to simply use multiple TCP connections for multiple application streams, but this also means

Figure 1. An inefficient choice of packet sizes

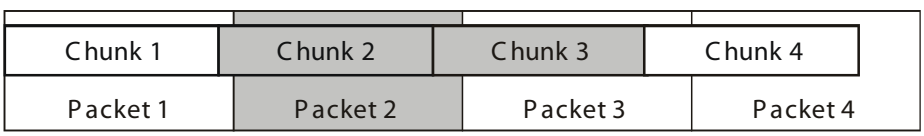
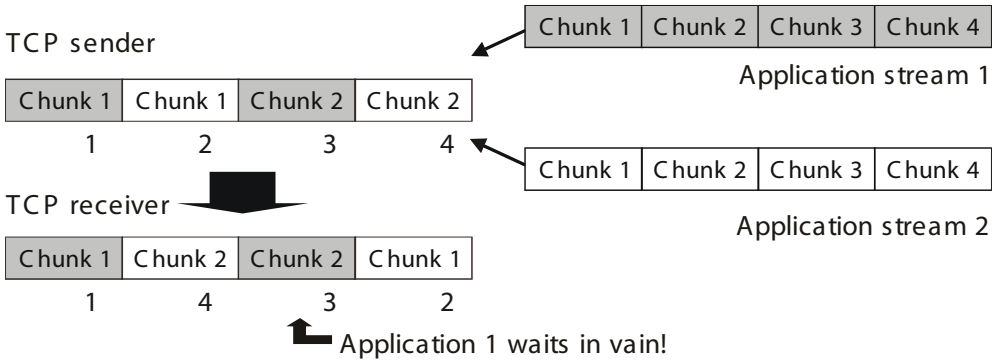


Figure 2. Transmitting two data streams over one TCP connection



that connection setup and teardown are carried out several times (thereby adding network traffic and increasing delay), and that congestion control is independently executed for each connection, rendering the total behavior of the source more aggressive than it should be (Balakrishnan, Rahul, & Seshan, 1999).

Multihoming

While TCP connections are uniquely identified via two IP addresses and two port numbers, SCTP connections are identified via two sets of IP addresses and two port numbers, and they are actually called "associations" instead of "connections." Multihoming at the transport layer is a powerful concept; it can enable an application to switch from one IP address to another when the communication fails without even noticing it. From the perspective of an application, the transport layer simply becomes more robust when multiple IP addresses are used for an association endpoint. The possible failure is not limited to the machine at the other end—SCTP can also switch when the communication flow is interrupted because of a problem inside the network. This can be used to shorten the time it takes for the network to "repair" an error (e.g., bypassing a failed link—since routing updates are typically sent every 30 seconds, the convergence time of Internet routing protocols can be quite long); SCTP can switch to an alternate address in the meantime and switch back when the problem has been solved. Multihoming may be particularly useful for mobile applications of any kind, where significant handover delays are known to be a common problem.

Partial Reliability

This feature, which was recently added to SCTP in a separate document (Stewart, Ramalho, Xie, Tuexen, & Conrad, 2004), makes it possible for an application to specify how persistent the protocol should be in attempting to deliver a message, including totally unreliable data transfer. This allows for multiplexing of unreliable and reliable data streams across the same connection; the ability to unreliably transfer data with congestion control functionality in place makes the service provided by this usage mode of SCTP quite similar to DCCP with TCP-like behavior (we will get to that later in this chapter), but with the additional benefit of features like multihoming.

UDP-LITE

If we regard SCTP as "TCP++," then UDP-Lite is "UDP++"—or actually "UDP--"—because its only feature is the possibility to restrain or even disable the original UDP checksum (Larzon et al.,

2004). The reason to do so is easily explained: there are video and audio codecs that can deal with bit errors (which can, for example, be caused by link noise in a wireless environment). However, even if only a single bit is wrong, the UDP checksum will fail, causing the receiver to drop the whole packet from the stack. The codec then ends up with a large number of bytes missing, as potentially useful data that actually made it to the receiver were discarded by the operating system.

The UDP-Lite header is very similar to the UDP header—just the "Length" field, which is redundant because the length of a datagram is contained in the IP header, was replaced with a field called "Checksum Coverage." It represents the number of bytes, counting from the beginning of the UDP-Lite header, that are covered by the checksum. Such partial coverage can be useful for certain codecs—the "adaptive multi-rate" and "adaptive multi-rate wideband" audio codecs, for example (Sjoberg, Westerlund, Lakaniemi, & Xie, 2002). In any case, it is mandatory to have the header checked because, without knowing that the header is correct, even the port numbers can be wrong and the whole communication flow becomes meaningless (it is actually possible to disable the checksum altogether in standard UDP, but this feature is rather useless).

Despite its simplicity and its seemingly obvious advantages, UDP-Lite caused a lot of discussions in the IETF. The main problem is the fact that UDP-Lite does not yield any benefits whatsoever unless a link layer technology actually hands over corrupt data. Since it is the first IETF development to have that requirement, link layer technologies were so far optimized for protocols that require data integrity. Typically, there is a strong checksum, and often, corrupt frames are retransmitted with a certain persistence and eventually dropped and not forwarded by the link layer (Fairhurst & Wood, 2002); this is, for example, the case with standard 802.11 wireless LAN systems.[2] UDP-Lite can be seen as being at odds with the notion "IP over everything," as it enables application programmers to write an application that works well in one environment (where there is a small loss ratio) and does not work at all in another. These issues are actually quite intricate; more details can be found in Welzl (2005). In any case, from the perspective of a mobile multimedia application programmer, UDP-Lite is probably an attractive protocol, and after a couple of years of discussion, it has been published as a "Proposed Standard" RFC by the IETF. Since it was designed to be downward compatible with UDP, there is not much harm in using it even though the benefits can only be attained if an underlying link layer hands over corrupt data.

THE DATAGRAM CONGESTION CONTROL PROTOCOL (DCCP)

Multimedia applications are supposed to adapt their rate to the allowed transmission rate of the network in order to prevent Internet congestion collapse; ideally, this should be done in a way that is fair towards TCP ("TCP-friendly"). This is easier said than done: simply using TCP is usually not an option, as most real-time multimedia applications put timely delivery before reliability (i.e., users normally accept some noise in a telephone conversation, but having the sound excessively delayed is intolerable). Thus, such applications use UDP instead of TCP—but UDP provides no congestion control, leaving the task up to its user.

Adapting the rate is generally a difficult issue at the application level: data can be layered, compression factors can be tuned, encodings can be changed, but the outcome is not always precisely predictable. The additional requirement of being fair towards TCP and embedding a complete congestion control mechanism within the application may just be too much for most developers. Moreover, there is an incentive problem here—while TCP-friendliness is definitely desirable from the network point of view, it is questionable whether

implementing this functionality is worth the effort for a single multimedia application developer. Finally, a user space congestion control implementation is just not ideal because precise timing may be necessary.

The IETF seeks to counter these problems with the datagram congestion control protocol (DCCP), which embodies congestion control functions for applications that do not need reliability. This protocol should be used as a replacement for UDP by networked multimedia applications and could be regarded as a framework for TCP-friendly mechanisms; due to a wealth of additional functions, DCCP is indeed an attractive alternative. According to its main specification (Kohler et al., 2005), one way of looking at the protocol is as TCP minus byte stream semantics and reliability, or as UDP plus congestion control, handshakes, and acknowledgments—but in fact, DCCP, includes much more than these three functions. In what follows, we will take a closer look at the most important elements of the protocol.

Connection Handling

Despite being unreliable, DCCP is connection oriented. The main reason for embedding this function in the protocol is to facilitate traversal of middle boxes such as firewalls which can selectively admit or reject communication flows when packets are associated with connections.

Reliable ACKs

Congestion control requires feedback. As in TCP, this takes the form of acknowledgment packets (ACKs) in DCCP—but with different semantics. In TCP, "ACK 2000" means "I received everything up to byte 1999, and I would like to have byte number 2000". Since DCCP never retransmits a packet, such cumulative ACKs would not make much sense here; thus, DCCP ACKs only acknowledge the reception of individual packets. This means that a sender has to maintain state regarding all the packets that were ACKed, and that it is hard for a sender to decide when to remove the state (at a TCP sender, the reception of a cumulative ACK can be used as an indication to remove any state regarding previous packets). It was therefore decided to make ACKs reliable in DCCP, in other words, retransmit them until the ACKs themselves are successfully ACKed. This has the additional advantage that congestion control can be carried out along the backward path—something that is hard to achieve with TCP, where data packets are reliably transferred, but ACKs are not. This fact makes DCCP superior to TCP in highly asymmetric environments such as satellite access links, where incoming data are streamed across a satellite and ACKs are often sent across a low-bandwidth modem link.

Feature Negotiation

In the DCCP specification, the word "feature" refers to a variable which is used to identify whether a DCCP endpoint uses a certain function. The "congestion control ID (CCID)" is an example of such a feature—endpoints must be able to negotiate which congestion control mechanism is to be used. The specification describes exactly how this is done; this includes the possibility to specify a "preference list," which is like saying "I would like CCID 2, but otherwise, please use CCID 1. If you cannot even give me CCID 1, let us use CCID 3." This procedure is another example of a reliable process that is embedded in this otherwise unreliable protocol. Features are specific to one endpoint, which means that a full-duplex DCCP communication flow can use one congestion control mechanism in one direction and another one in the other.

Checksums

As with UDP-Lite, the DCCP checksum can be restricted or even completely disabled. Additionally, the "data checksum" option can be used to

distinguish between corruption-based loss and other loss events even when it is unacceptable to deliver erroneous data to applications. With this mechanism, a DCCP congestion control mechanism can therefore bypass the well-known TCP problem of misinterpreting any kind of packet loss as a sign of congestion (Balan et al., 2001).

Full Duplex Communication

Applications such as VoIP or video conferencing tools may require a bidirectional data stream—here, the full duplex communication capability of DCCP can make things more efficient by piggybacking ACKs onto data packets. Additionally, having only one logical connection for two unidirectional flows facilitates middle box traversal (i.e., firewalls require less state—and making the job easy for firewall developers fosters deployment of the protocol).

Explicit Congestion Notification (ECN) Support

With ECN, routers are given the option to set a bit in packets that they would normally drop (Ramakrishnan et al., 2001); the underlying idea is that a receiver should inform a sender when the "congestion experienced (CE)" bit in the IP header was set by a router, and a sender should react as if the packet had been dropped. With UDP, where the application programmer is responsible of implementing proper congestion control, ECN support could lead to unfairness—after all, who could prevent an application programmer from simply ignoring the bit? Thus, the fact that DCCP makes use of ECN could be seen as a function that makes it somewhat superior over UDP.

Security

DCCP was designed to be at least as secure as a state of the art TCP implementation; modern TCP functions like ECN Nonces (a mechanism that prevents a receiver from lying about the congestion state) (Spring, Wetherall, Ely, 2003) and "appropriate byte counting (ABC)" (Allman, 2003) as well as "cookies" that can reduce the chance for a TCP-SYN-like DoS attack to succeed are therefore part of the protocol.[3] In TCP, sequence numbers automatically yield some protection against hijacking attacks; due to its unreliable nature, this had to be taken care of by means of a special sequence number synchronization procedure in DCCP.

Mobility

Whether to support mobility or not was discussed at length in the DCCP working group; eventually, neither mobility nor multihoming were included in the main document, and the specification was postponed. A rudimentary mechanism that slightly diverges from the original DCCP design rationale of not using cryptography is currently in the works. It is disabled by default, and an endpoint that wants to use this mechanism must negotiate enabling the corresponding feature. The scheme is simpler than mobility support in SCTP and resembles Mobile IP as specified in RFC 3344 (Perkins, 2002); at this point in time, it is unclear if (or when) it will be published as an RFC.

CONCLUDING REMARKS

The sudden appearance of new transport protocols for the Internet may help to make things more efficient, but it certainly will not make them easier to handle. How is an application programmer supposed to know whether, say, SCTP with parameters chosen for unreliable transmission or DCCP with TCP-like behaviour is more suitable for a certain situation? Also, these new transport protocols may face severe deployment problems—there must be a clear incentive for an application programmer to use a new protocol, which always has the potential risk of not

penetrating an outdated firewall. Requiring to update all DCCP-based applications whenever a new CCID becomes defined also does not seem to be very attractive. Many more questions appear on the horizon—for instance, how do RTP and DCCP go together?[4] Finally, experiences with these protocols in mobile environments is quite limited.

While development of these protocols has progressed nicely and already reached a certain level of maturity, their use is still in its infancy. It may take a while until the radical change from TCP and UDP to a total of five transport protocols is welcomed by the majority of application developers; in any case, at this point in time, putting some research efforts into studying their usage in different scenarios seems to be a good idea.

REFERENCES

Allman, M. (2003). *TCP congestion control with appropriate byte counting (ABC)* (Tech. Rep. No. RFC 3465). Internet Engineering Task Force (IETF).

Balakrishnan, H., Rahul, H. S., & Seshan, S. (1999). An integrated congestion management architecture for Internet hosts. In *Proceedings of SIGCOMM,* Cambridge, MA (pp. 175-187).

Balan, R. K., Lee, B. P., Kumar, K. R. R., Jacob, J., Seah, W. K. G., & Ananda, A. L. (2001). TCP HACK: TCP header checksum option to improve performance over lossy links. *20th IEEE Conference on Computer Communications (INFOCOM).*

Clark, D., & Tennenhouse, D. (1990). Architectural considerations for a new generation of protocols. In *Proceedings of SIGCOMM,* Philadelphia (pp. 200-208).

Fairhurst, G., & Wood, L. (2002). *Advice to link designers on link* (Tech. Rep. No. RFC 3366). Automatic Repeat reQuest (ARQ).

Floyd, S., & Fall, K. (1999). Promoting the use of end-to-end congestion control in the Internet. *IEEE/ACM Transactions on Networking, 7*(4), 458-472.

Floyd, S., Handley, M., Padhye, J., & Widmer, J. (2000). Equation-based congestion control for unicast applications. In *Proceedings of ACM SIGCOMM,* Stockholm, Sweden (pp. 43-56).

Hessler, S., & Welzl, M. (2005). An empirical study of the congestion response of RealPlayer, Windows MediaPlayer, and Quicktime. In *Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC),* La Manga del Mar Menor, Cartagena, Spain.

Kohler, H., Handley, M., & Floyd, S. (2005). *Datagram Congestion Control Protocol (DCCP).* Internet-draft draft-ietf-dccp-spec-11.txt. Retrieved from http://www.icir.org/kohler/dccp/

Larzon, L. A. , Degermark, M., Pink, S., Jonsson, L. E. & Fairhurst, G. (2004). *The lightweight user datagram protocol (UDP-Lite)* (Tech. Rep. No. RFC 3828). Internet Engineering Task Force (IETF).

Perkins, C. (2002). *IP mobility support for IPv4* (Tech. Rep. No. RFC 3344). Internet Engineering Task Force (IETF).

Ramakrishnan, K., Floyd, S., & Black, D. (2001). *The addition of explicit congestion notification (ECN) to IP* (Tech. Rep. No. RFC 3168). Internet Engineering Task Force (IETF).

Sjoberg, J., Westerlund, M., Lakaniemi, A., & Xie., Q. (2002). *Real-time transport protocol (RTP) payload format and file storage format for the adaptive multi-rate (AMR) and adaptive multi-rate wideband (AMR-WB) audio codecs* (Tech. Rep. No. RFC 3267). Internet Engineering Task Force (IETF).

Spring, N., Wetherall, D., & Ely, D. (2003). *Robust explicit congestion notification (ECN) signaling with nonces* (Tech. Rep. No. RFC 3540). Internet Engineering Task Force (IETF).

Stewart, R., Ramalho, M., Xie, Q., Tuexen, M., & Conrad, P. (2004). *Stream control transmission protocol (SCTP) partial reliability extension* (Tech. Rep. No. RFC 3758). Internet Engineering Task Force (IETF).

Stewart, R., & Xie, Q. (2002). *Stream control transmission protocol (SCTP). A reference guide*. Boston: Addison-Wesley.

Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L., & Paxson, V. (2000). *Stream control transmission protocol* (Tech. Rep. No. RFC 2960). Internet Engineering Task Force (IETF).

Welzl, M. (2005). Passing corrupt data across network layers: An overview of recent developments and issues. *EURASIP Journal on Applied Signal Processing, 2005*(2), 242-247.

KEY TERMS

Application Layer Framing (ALF): Putting an application in control of block sizes that are transferred across the network.

DCCP: Datagram congestion control protocol.

Head-of-Line Blocking Delay: Delay that is caused by the requirement to deliver data chunks in order.

Multihoming: Associating a single logical connection endpoint with multiple IP addresses.

SCTP: Stream control transmission protocol

ENDNOTES

1. Internet Engineering Task Force–the technical standardization body of the Internet.
2. WiMAX (802.16) is a counter-example: here, it is possible to disable the checksum, albeit for reasons of compatibility with ATM.
3. Cookies can also be found in the SCTP association setup procedure (Stewart & Xie, 2002).
4. The answer to this question is: while DCCP functions could theoretically be implemented on top of RTP, it was decided that having RTP run over DCCP would be the right way to proceed.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 129-138, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.5
Future Directions of Multimedia Technologies in E-Learning

Timothy K. Shih
Tamkang University, Taiwan

Qing Li
City University of Hong Kong, Hong Kong

Jason C. Hung
Northern Taiwan Institute of Science and Technology, Taiwan

ABSTRACT

In the last chapter, we discuss how advanced multimedia technologies are used in distance learning systems, including multimedia authoring and presentation, Web-based learning, virtual environments, interactive video, and systems on mobile devices. On the other hand, we believe pedagogic theory should be incorporated into the design of distance learning systems to add learning efficiency. Thus, we point out some suggestions to the designers of future distance learning systems.

INTRODUCTION

Distance learning, based on styles of communication, can be categorized into synchronized and asynchronized modes. The advantages of distance learning include flexibility of time and space, timely delivery of precisely presented materials, large amount of participants and business opportunity, and automatic/individualized lecturing to some degrees. Both synchronized and asynchronized distance learning systems rely on multimedia and communication technologies. Due to its commercial value, distance learning is becoming a killer application of multimedia and communication research. We discuss current distance learning systems based on the types of

multimedia technologies used and point out a few new research directions in the last section.

MULTIMEDIA PRESENTATIONS AND INTERACTIONS

Authoring and playback of multimedia presentations are among the earliest applications of multimedia technologies. Before real-time communication and video-on-demand technologies, multimedia presentations were delivered to kids and distance learning students on CD ROMs. The advantage of multimedia presentation over traditional video tapes is due to interactivity. Multimedia presentations allow one to select "hot spots" in individualized topology. Techniques to realize this type of CD ROM presentations allow a rich set of media coding and playback mechanisms, such as images, sounds, and animations (including video and motion graphics). Successful examples include MS PowerPoint, Authorware Professional, Flash, and others.

With the development of communication technologies, multimedia computing focuses on efficient coding mechanism to reduce the amount of bits in transmission. Synchronization among media became important. Inner stream synchronization is implemented in a single multimedia record, such as the interleaving coding mechanism used in a video file, which includes sound track and motion picture track. Another example of inner stream synchronization and coding is to merge graphics animation with video stream (Hsu, Liao, Liu, & Shih, 2004). On the other hand, inter stream synchronization is more complicated since both the client (i.e., user) side and the server (i.e., management system) side need to work together. Inter stream synchronization allows packages (e.g., sound and image) to be delivered on different paths on a network topology. On the client side, packages are re-assembled and ensured to be synchronized. Another example of a recent practical usage of inter stream synchronization is in several commercial systems allowing video recording to be synchronized with MS PowerPoint presentation or Flash. Some systems (Shih, Wang, Liao, & Chuang, 2004) use an underlying technology known as the advanced streaming format (ASF) of Microsoft. ASF allows users or programs to embed event markers. In a playback system on the client side, users can interrupt a vide presentation, or jump to another presentation section. The video presentation can also use markers to trigger another presentation object such as to bring up a PowerPoint slide (converted to an image) or another multimedia reference. In order to deliver a synchronized presentation, an ASF server needs to be installed on the server machine.

ASF provides a preliminary technology for video-on-demand (or lecture-on-demand). In order to support multiple clients, it is necessary to consider bandwidth allocation and storage placement of video records. Video-on-demand systems (Hua, Tantaoui, & Tavanapong, 2004; Mundur, Simon, & Sood, 2004 allow a video stream to be duplicated and broadcast in different topology on multiple channels, to support multiple real-time requests in different time slots. In addition, adaptive coding and transmission mechanism can be applied to video-on-demand systems to enhance overall system performance.

Video-on-demand allows user interactions to select video programs, perform VCR-like functions, and choose language options. Interactive TV (Liao, Chang, Hsu, & Shih, 2005) further extends interactivities to another dimension. For instance, the users can select the outcome of a drama, refer to specification of a commercial product, or answer questions pre-defined by an instructor. The authoring and playback system developed in Liao, Chang, Hsu, and Shih (2005) takes a further step to integrate video browser (for interactive TV) and Web browser. Thus, distance learning can be implemented on set-top box.

WEB-BASED DISTANCE LEARNING AND SCORM

Most multimedia presentations can be delivered online over Internet. And, Web browser is a common interface. HTML, XML, and SMIL are the representation languages of learning materials. Typically, HTML is used in the layout while other programming languages (such as ASP) can be used with HTML to retrieve dynamic objects. As an extension to HTML, XML allows user defined tags. The advantage of XML allows customized presentations for different Web applications, such as music and chemistry, which requires different presentation vocabularies. In addition, SMIL incorporates controls for media synchronization in a relatively high level, as compared to inner stream coding technologies. HTML-like presentations can be delivered by Web servers, such as Apache and MS IIS.

Although Web browsers are available on different operating systems and HTML-like presentations can be reused, search and reuse of course materials, as well as their efficient delivery, are key issues to the success of distance learning. In order to achieve reusability and interoperability, a standard is needed. The advanced distributed learning (ADL) initiative proposed the sharable content object reference model (SCORM) (The Sharable Content Object Reference Model, 2004) standard since 2000. Main contributors to SCORM include the IMS Global Learning Consortium, Inc., the Aviation Industry CBT (computer-based training) Committee (AICC), the Alliance of Remote Instructional Authoring & Distribution Networks for Europe (ARIADNE), and the Institute of Electrical and Electronics Engineers (IEEE) Learning Technology Standards Committee (LTSC). The SCORM 2004 (also known as SCORM 1.3) specification consists of three major parts

- **The content aggregation model (CAM):** Learning objects are divided into three categories (i.e., assets, sharable content objects (SCOs) and content organizations). The contents of the learning objects are described by metadata. In addition, CAM includes a definition of how reusable learning objects are packed and delivered.
- **The run-time environment:** In order to deliver learning objects to different platforms, a standard method of communication between the learning management system (LMS) and the learning objects is defined.
- **The sequencing and navigation:** Interactions between users (i.e., students) and the LMS are controlled and tracked by the sequencing and navigation definitions. This also serves as a standard for defining learner profiles, as well as a possible definition for intelligent tutoring.

The SCORM specification clearly defines representation and communication needs of distance learning. To realize and promote the standard, a few SCORM-compliant systems were implemented (Chang, Chang, Keh, Shih, & Hung, 2005; Chang, Hsu, Smith, & Wang, 2005; Shih, Lin, Chang, & Huang; Shih, Liu, & Hsieh, 2003). However, common repository for SCORM learning objects, representation of learner records, and intelligent tutorial mechanisms to facilitate sequencing and navigation are yet to be identified. On the other hand, most existing SCORM-compliant LMSs fail to support the newest specification, except the prototype provided by ADL.

VIRTUAL CLASSROOM AND VIRTUAL LAB

Web-based distance learning supports asynchronized distance learning in general. Usually, distance learning programs rely on Web browsers to deliver contents, collect assignments from students, and allow discussion using chat room or e-mails. These functions can be integrated in a

distance learning software platform such as Blackboard (http://www.blackboard.com/) and WebCT (http://www.Webct.com/). On the other hand, real-time instruction delivery can be broadcast using video channels, or through bi-directional video conferencing tools (Deshpande, & Hwang, 2001; Gemmell, Zitnick, Kang, Toyama, & Seitz, 2000). Real-time video communication requires sophisticated network facilities and protocols to guarantee bandwidth for smooth transmission.

In addition to online delivery of instruction, lab experiments can be realized using remote labs or virtual labs (Auer, Pester, Ursutiu, & Samoila, 2003). Remote lab uses camera and advanced control technologies to allow physical lab instruments to be accessed by students using Internet. Virtual lab may or may not include physical experimental instruments. Emulation models are usually used. In most cases, assessment of experiment outcomes from software emulation is compared with those from physical devices.

Virtual reality (VR) and augmented reality techniques can also be used in distance learning (McBride & McMullen, 1996; Shih, Chang, Hsu, Wang, & Chen, 2004). Most VR systems use VRML, which is an extension of XML for 3-D object representations. The shared-Web VR system (Shih, Chang, Hsu, Wang, & Chen, 2004) implements a virtual campus, which allows students to navigate in a 3-D campus, with different learning scenarios. Behaviors of students can be tracked and analyzed. The incorporation of game technologies points out a new direction of distance learning, especially for the design of courseware for kids. With wireless communication devices, ubiquitous game technologies can be used for mobile learning in the near future.

MOBILE LEARNING

Wireless communication enables mobile learning. With the capability of multimedia technologies on wireless connected notebook computers, PDAs, and even cellular phones, system developers are possible to implement distance learning systems on mobile devices (Meng, Chu, & Zhang, 2004; Shih, Lin, Chang, & Huang, 2004). The challenges of deploying course materials on small devices, such as cellular phones, include the limited display space, slow computation, and limited memory capacity. On a small display device, reflow mechanism can be implemented (Shih, Lin, Chang, & Huang, 2004). The mechanism resizes contents into a single column layout, which can be controlled using a single scroll bar on PDAs or cellular phones. To cope with small storage, prefetching technique on subdivided course contents can be used. Thus, the readers can download only the portion of contents of interesting.

To realize learning management systems on wireless network connected devices, a distributed architecture needs to be designed between the server and the client (e.g., PDA). SOAP is a communication protocol very suitable for the architecture. SOAP packages are messages that can be sent between a client and server, with a standard representation envelope recommended by the W3C (http://www.w3.org/). The advantages of the protocol include platform independency, accessibility, and implementation efficiency. In addition, in order to maintain the status of each individual learner, learner profiles needs to needs to be defined. Yet, SCORM contains only a preliminary description of learner profile definition. The representation of course contents should also consider how to enable small packages to be delivered on a remote request. Cashing mechanism and hand shaking protocol are important issues yet to be developed. In addition, in some occasion for situated learning, location awareness is necessary for situated collaborative learning.

On the other hand, synchronized distance learning on mobile devices requires efficient real-time streaming due to the limited bandwidth of current wireless communication systems (Liu, Chekuri, & Choudary, 2004. Even as 3G mobile communication technologies are avail-

able, smooth video streaming requires a broader channel and a robust error resilience transmission mechanism.

HYBRID INTERACTIVE SYSTEMS AND PEDAGOGICAL ISSUES

Whether learning activities are implemented on mobile devices or PC clients, efficient collaboration is the key issue toward the success of learning. A SCORM-based collaborative learning LMS is developed in Chang, Lin, Shih, and Wang (2005). The system allows learning activities among students to be synchronized based on the Petri net model. The instructor is able to supervise the collaboration behavior among a group of students. Whether or not it is SCORM compliant, a distance learning platform should support collaboration in either synchronized or asynchronized manner. At least, a CSCW-like system should be implemented to support the need of collaboration.

Recently, personalized Web information delivery has become an interesting issue in data mining research. A distance learning server is able to analyze student profiles, depending on individual behaviors. Learner profiles can be stored and analyzed according to traversal sequences and results from tests. This type of distance learning system is based on the self-regulation principles of the social cognitive theory (Bandura, 1986). A system using this approach should allow students to plan on their study schedule based on individual performance (Leung, & Li, 2003), while the underlying intelligent mechanism can guide students to a suitable study schedule, which can be reviewed by an instructor. To facilitate user friendliness, self-regulation can be incorporated with Web-based interfaces and mobile devices. To some degree of the usage of artificial intelligence (Shih & Davis, 1997), an intelligent tutorial system is able to generate individualized lectures (Leung & Li, 2003).

We realize that, it is possible to design an integrated learning environment to support the application of the scaffolding theory (Zimmerman & Schuck, 1989). Scaffolding, proposed by L. S. Vygotsky, was viewed as social constructivism. The theory suggests that students take the leading role in the learning process. Instructors provide necessary materials and support. And, students construct their own understanding and take the major responsibility. Between the real level of development and the potential level of development, there exists a zone of proximal development. This zone can be regarded as an area where scaffolds are needed to promote learning. Scaffolds to be provided include vertical and horizontal levels as a temporary support in the zone of proximal development. The scaffolding theory is essential for cognitive development. It also supports the process of social negotiation to self-regulation. There are three properties of the scaffold

- The scaffold is a temporary support to ensure the success of a learning activity.
- The scaffold is extensible (i.e., can be applied to other knowledge domains) and can be used through interactions between the learner and the learning environment.
- The scaffold should be removed in time after the learner is able to carry out the learning activities independently.

The scaffolding theory indicates three key concepts. Firstly, in the zone of proximal development, the relationship between the scaffolds providers and the receivers are reciprocal. That means that the instructor and students negotiate a mutual beneficial interactive process. Secondly, the responsibility is transferred from the instructor to the student during the learning process. Depending on the learning performance, the instructor gradually gives more control of the learning activities to the student for the ultimate goal of self-regulation. Finally, the interaction facilitates the learners to organize their own

knowledge. Scaffolding also encourages the use of language or discourse to promote reflection and higher-order thinking.

Pedagogical principles are not multimedia technology. However, the developers of distance learning system should be aware of the concept.

SUMMARY

This chapter summarizes multimedia technologies for distance learning systems. While we were looking for the essential needs of professional educators and students, in terms of "the useful multimedia distance learning tools," we have found that lots of tools were developed by computer scientists. Most of these tools lack of underlying educational theory to show their usability. However, software is built for people to use. In spite of its advanced functionality and outstanding performance, any system will be useless if no one uses it. Thus, we believe the specification of a distance learning system should be written by educational professionals, with the help of computer scientists.

From the perspective of multimedia and Internet computing, there are a few challenging research issues to make distance-learning systems more colorful and useful. We highlight a few here

- **Interactive TV:** Video-on-demand technologies should be highly integrated with interactive TV and set-top box devices, which should be extended to incorporate different modals of interaction. A sophisticated bi-directional inter stream synchronization mechanism needs to be developed.
- **Standards:** The most popular standard is SCORM. However, the definition of user profile, federal repository, and adaptive techniques for mobile devices are yet to be investigated.
- **High communication awareness:** Video conferencing tools should be integrated with awareness sensors, to bring the attentions on interested video area to users.
- **Virtual and remote lab:** A standard development specification for creating virtual or remote labs is not yet developed. The standard should allow reusable lab components which can be assembled to facilitate different varieties of lab designs.
- **Adaptive contents for mobile learning:** Different mobile devices should have different functional specifications to guide a central server to transmit device and user dependent media for efficient learning.
- **Intelligent tutoring:** User profile dependent tutoring based on intelligent technology applied on Web technology should be used. Pedagogical considerations can be applied on intelligent tutoring.

Among the developed platforms for distance learning, an assessment mechanism, especially the one based on educational perspective, should also be proposed. It is the hope that the multimedia research community can work with educational professionals and the distance learning industry together, to develop a standard distance-learning framework for the success of our future education.

REFERENCES

Auer, M., Pester, A., Ursutiu, D., & Samoila, C. (2003, December). Distributed virtual and remote labs in engineering. *International Conference on Industrial Technology ICIT 2003,* Slovenia (pp. 1208-1213).

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Chang, F. C., Chang, W., Keh, H., Shih, T. K., & Hung, L. (2005). Design and implementation of a SCORM-based courseware system using influence diagram. *International Journal of Distance Education Technologies, 3*(3), 82-96.

Chang, W., Hsu, H., Smith, T. K., & Wang, C. (2005). Enhancing SCORM metadata for assessment authoring in e-learning. *Journal of Computer Assisted Learning, 20*(4), 305-316.

Chang, W., Lin, H. W., Shih, T. K., & Wang, C. (2005, March 28-30). Applying Petri nets to model SCORM learning sequence specification in collaborative learning. *Proceedings of the 19th International Conference on Advanced Information Networking and Applications*, Taiwan.

Deshpande, S. G., & Hwang, J. (2001, December). A real-time interactive virtual classroom multimedia distance learning system. *IEEE Transactions on Multimedia, 3*(4), 432-444.

Gemmell, J., Zitnick, L., Kang, T., Toyama, K., & Seitz, S. (2000, October-December). Gaze-awareness for video conferencing: A software approach. *IEEE MultiMedia, 7*(4), 26-35.

Hsu, H. H., Liao, Y. C., Liu, Y.-J., & Shih, T. K. (2004). Video presentation model. In S. Deb (Ed.), *Video data management and information retrieval* (pp. 177-192). Hershey, PA: Idea Group Publishing.

Kien, A., Hua, M. A., Tantaoui, & Tavanapong, W. (2004, September). Video delivery technologies for large-scale deployment of multimedia applications. *Proceedings of the IEEE* (Special issue on Evolution of Internet Technologies towards the Business Environment), *92*(9), 1439-1451.

Leung, E. W. C., & Li, Q. (2002). Media-on-demand for agent-based collaborative tutoring systems on the Web. *IEEE Pacific Rim Conference on Multimedia 2002* (pp. 976-984).

Leung, E. W. C., & Li, Q. (2003). A dynamic conceptual network mechanism for personalized study plan generation. *ICWL 2003* (pp. 69-80).

Liao, Y., Chang, H., Hsu, H., & Shih, T. K. (2005). Merging web browser and interactive video: A hypervideo system for e-learning and e-entertainment. *Journal of Internet Technology, 6*(1), 121-131.

Liu, T., & Choudary, C. (2004, October). Realtime content analysis and adaptive transmission of lecture videos for mobile applications. *Proceedings of the 12th ACM International Conference on Multimedia*, New York.

McBride, J. A., & McMullen, J. F. (1996, January). Using virtual reality for distance teaching a graduate information systems course. *Proceedings of the 29th Hawaii International Conference on System Sciences 1996* (Vol. 3, pp. 263-272).

Meng, Z., Chu, J., & Zhang, L. (2004, May). Collaborative learning system based on wireless mobile equipments. *IEEE Canadian Conference on Electrical and Computer Engineering CCECE 2004* (Vol. 1, pp. 481-484).

Mundur, P., Simon, R., & Sood, A. (2004, February). End-to-end analysis of distributed video-on-demand systems. *IEEE Transactions on Multimedia, 6*(1), 129-141.

The Sharable Content Object Reference Model. (2004). *ADL Co-Laboratory.* Retrieved from http://www.adlnet.org/

Shih, T. K., Chang, Y., Hsu, H., Wang, Y., & Chen, Y. (2004). A VR-based shared Web system for distance education. *International Journal of Interactive Technology and Smart Education (ITSE), 1*(4), 4.

Shih, T. K., & Davis, R. E. (1997, April-June). IMMPS: A multimedia presentation design system. *IEEE Multimedia* (pp. 67-78).

Shih, T. K., Lin, N. H., Chang, H., & Huang, K. (2004, June 27-30). Adaptive pocket SCORM reader. *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME2004),* Taipei, Taiwan.

Shih, T. K., Liu, Y., & Hsieh, K. (2003, July 6-9). A SCORM-based multimedia presentation and editing system. *Proceedings of the 2003 IEEE International Conference on Multimedia & Expo (ICME2003),* Baltimore.

Shih, T. K., Wang, T., Liao, I., & Chuang, J. (2003). Video presentation recording and online broadcasting. *Journal of Interconnection Networks* (Special issue on Advanced Information Networking: Architectures and Algorithms), *4*(2), 199-209.

Zimmerman, B. J., & Schuck, D. H. (1989). *Self-regulated learning and academic achievement.* New York: Springer-Verlag.

This work was previously published in Future Directions in Distance Learning and Communication Technologies, edited by T. K. Shih, pp. 273-283, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 8.6
Toward Effective Use of Multimedia Technologies in Education

Geraldine Torrisi-Steele
Griffith University, Australia

ABSTRACT

While multimedia technologies are being used in educational contexts, the effective use of multimedia in these contexts remains problematic. In an attempt to contribute towards addressing this problem, this chapter presents a set of conceptual guidelines and a practical planning framework that is intended to inform the planning and design of more effective multimedia integration into educational contexts. A mixed-mode approach is advocated in this chapter. Multimedia technologies are viewed as part of a tool-set and tool selection should be appropriate to curriculum content and to the teaching and learning context.

INTRODUCTION

Whether or not multimedia technologies should be used in educational contexts seems to no longer be an issue. Multimedia technology is pervading almost all aspects of existence. The rationale for its use in educational contexts is grounded in social, economic, and pedagogical reasons. However, what does remain problematic is the effective use of multimedia technology in educational contexts. At the crux of addressing this problem is the notion that effective integration of multimedia in the curriculum depends not on the technology itself but rather on educators' knowledge, assumptions, and perceptions regarding the technology and its implementation in the specific learning context (Jackson & Anagnotopoulou, 2000; Bennet, Priest, & Macpherson, 1999). From a pedagogical perspective, it is generally accepted that multimedia technologies have the potential to reshape and add a new dimension to learning (Relan & Gillani, 1997; Lefoe, 1998). In reality, however, this potential has largely failed to be realized. The fundamental belief underlying this chapter is that this potential will only be realized by informed pedagogical decision making and the formulation of teaching strategies designed to

exploit multimedia technologies for maximum effectiveness within a particular learning situation. From this perspective, educator development that focuses on pedagogical change is a pivotal aspect of the effective use of multimedia technologies in educational contexts.

The term "multimedia technologies" is being used in this chapter to mean the entirely digital delivery of content using any integrated combination of audio, video, images (two-dimensional, three-dimensional), and text. In its most primitive form, the term "multimedia" is sometimes defined as content presentation using a combination of media [i.e., sound, images (static, moving, animated, video), and text]. From this perspective, any presentation that involves the use of, for example, face-to-face teaching, video recorder, and a slide show could be considered multimedia.

The distinguishing feature of digital multimedia, as used in this chapter (as opposed to the primitive form defined above), is the capacity to support user interaction. Hence, the term "multimedia technologies," as used in this chapter, will always imply that there is an element of "interactivity" present. The concept of interaction is considered along two dimensions: the capacity of the system to allow an individual to control the pace of presentation and to make choices about which pathways are followed to move through the content, and the ability of the system to accept input from the user and provide appropriate feedback to that input. Multimedia technologies may be delivered on computer via CD-ROM, DVD, or via the Internet, or on other devices such as mobile phones and personal digital assistants capable of supporting interactive and integrated delivery of digital audio, video, image, and text data. Multimedia technologies as referred to in this chapter also encompass new communications technologies such as e-mail, chat, and videoconferencing. Virtual reality technologies are also included.

It will be argued later in this chapter that various multimedia technologies are seen as part of a tool set or possible modes of instruction. Other modes include face-to-face teaching, print materials, and video and audio devices. A "mixed-mode" approach will be advocated in this chapter based on the argument that tool selection should be appropriate to curriculum content and to the teaching and learning context.

The contents of this chapter have been based largely on the author's professional development experiences with tertiary educators implementing online learning. However, the ideas discussed in this chapter are based on principles of good practice that apply to a broad range of teaching and learning contexts, including primary, secondary, tertiary, and other training environments.

Against this background, this chapter aims to provide a set of conceptual guidelines and a practical foundation (in the form of a planning framework) that will be of interest to those involved in planning and designing appropriate professional development targeted at promoting effective multimedia integration, and to individual educators in primary, secondary, tertiary, and other training environments who wish to implement multimedia technologies more effectively into the curriculum.

MULTIMEDIA TECHNOLOGIES IN LEARNING ENVIRONMENTS

When computer-based interactive multimedia emerged in the 1990s, innovative educators began considering what implications this new media might have if it was applied to teaching and learning environments. Within a relatively short time frame, the emerging multimedia and associated communications technologies infiltrated almost every aspect of society. So, what was initially viewed as a technology "option" in educational contexts has for social, economic, and pedagogical reasons become a "necessity." Many educational institutions are investing considerable time, effort, and money into the use of technology.

Socially, computer literacy is an essential skill for full participation in society. The use of multimedia technologies in educational institutions is seen as necessary for keeping education relevant to the 21st century (Selwyn & Gordard, 2003).

Economically, the belief prevails that the large-scale use of new multimedia and associated communication technologies for teaching and learning may offer cheaper delivery than traditional face-to-face and distance education and will also help establish and maintain competitive advantages for institutions by allowing them to tap into overseas markets (Bennet, Priest, & Macpherson, 1999, p. 207).

The pedagogical basis for the use of educational multimedia technologies has perhaps been the greatest driving force for the massive investments made by educational institutions into multimedia technologies. Literature abounds with rhetoric about the potential impact of multimedia technologies on traditional teaching practices. The central theme is that the integration of multimedia technologies will lead to a transformation of pedagogy from traditional instructivist teacher-centered approaches to the more desirable constructivist learner approaches that are seen as embodying essential characteristics of more effective learning environments (Tearle, Dillon, & Davis, 1999; Relan & Gillani, 1997; Willis & Dickson, 1997; LeFoe, 1998; Richards & Nason, 1999). From the learner-centered perspective, the teacher's role changes from the traditional (instructivist approach) role of instructor and supplier of knowledge to a role more closely aligned with support and facilitation of the active construction of knowledge by the learner (Tearle, Dillon, & Davis, 1999). The learner-centered approach implies empowerment of the individual learner and the ability to provide the learner with self-directed, more meaningful, authentic learning experiences that lead to lifelong learning. This implication is at the crux of constructivist-based pedagogical arguments for the integration of multimedia technologies in educational contexts (Selwyn & Gorard, 2003; Gonzales et al., 2002).

However, despite the well-documented and generally accepted potential of multimedia technologies to reshape teaching practices, it has been identified in literature that the promised impact of multimedia technologies on learning and pedagogical practices have largely not eventuated. There are relatively few positive impacts on educational practices for major investments of time, effort, and money by educational institutions (Cuban, 1986; Hammond, 1994; Oliver, 1999; Nichol & Watson, 2003; Conlon & Simpson, 2003; Selwyn & Gorard, 2003).

The reason for this lack of impact is seen to lie not with the attributes of the technology itself, but rather with the ways in which the technology has been implemented in learning contexts. More specifically, it is the educators' knowledge, assumptions, and perceptions regarding the technology and its implementation in the specific learning context that will determine its implementation and, hence, its effectiveness (Jackson & Anagnotopoulou, 2000; Bennet, Priest, & Macpherson, 1999). As is often noted in literature, the potential of multimedia technologies to reshape learning contexts (Relan & Gillani, 1997; Lefoe, 1998) will only be realized by informed pedagogical decision making and the formulation of teaching strategies designed to exploit multimedia technologies within the curriculum context.

Although it may be recognized by educators that multimedia technologies have the potential to offer new and improved learning opportunities, many educators fail to realize this potential. A number of educators using multimedia technologies in their learning environments are largely limiting its use to a tool for data access, communication, and administration (Conlon & Simpson, 2003). This is an "add-on" approach to multimedia technology use rather than a truly integrated curriculum approach. This lack of true integration results in minimal (if any) change in both pedagogical strategies and learning environment (Tearle, Dillon, & Davis, 1999; Strommen, 1999).

Failure to implement effective technology integration is attributed to the fact that educators, even experienced educators, are generally unprepared for the changes demanded by and produced by "technology infusion" (Charp, 2000). While some of the pedagogical "know how" of more traditional learning environments possessed by educators may transfer to new interactive multimedia contexts, educators often lack the skills and technical and pedagogical knowledge to effectively implement those technologies in their learning environments. Rakes and Casey (2002, online) observed the following:

…many [educators], especially more experienced teachers, have been unable to find effective ways to use technology in their classrooms. One possible explanation for this lack of success is that the use of technology in the classroom has been viewed in terms of simple skill acquisition instead of as a change process that affects the behavior of individuals on a very profound level.

If there is a lesson to be learned from the last few decades of "educational technology" development, it is that technologies themselves offer very little to the learning process. Conlon & Simpson (2003, p. 149) warned that if educators are "hastened" into adopting multimedia technologies without any clear educational vision of change, then significant transformation of teaching practice is unlikely. The importance of focus on educator development and resources that will foster continuous pedagogical growth and "re-engineering" becomes self-evident and is well documented in literature (Gonzales et al., 2002; Burns, 2002; Pierson, 2001; Charp, 2000; Collis, 1996; Rakes & Casey, 2002).

Against this context, some of the key issues that need be addressed in educator development will be identified and discussed. Five key guidelines and a planning framework for facilitating more effective multimedia technologies integration will be presented.

TOWARD MORE EFFECTIVE TECHNOLOGY INTEGRATION

The preceding discussion has directed attention to the notion that while multimedia technologies have the potential to reshape practice, the potential is often unrealized due to the fact that educators are often ill-equipped to meet the challenges of change demanded by multimedia technologies and to exploit change made possible by them. This notion is supported by an earlier study (Torrisi & Davis, 2000) conducted by the author into the experiences of tertiary educators developing online multimedia materials.

The data from the study highlight some of the key issues that need to be addressed in educator development efforts. Educators in the study were asked to identify what they believed were key competencies that students should develop as a result of undertaking study in the subject. Each educator was also asked to clarify what they believed to be the role of online materials in their course. Table 1 juxtaposes individual educator's responses for key competencies against the educator's stated intended use of online materials. Upon examination of responses as shown in Table 1, a lack of congruency between what educators identified as key competencies for their students and the stated use of online materials was found. This lack of congruency between stated key competencies and intended use of online materials is indicative of multimedia technology that is not truly integrated with the curriculum goals, content, objectives, and context, rather use is limited to being add-on or supplemental.

Insight into reason for supplemental use of multimedia technologies was revealed in interviews with the tertiary educators, whose comments suggested they perceived the use of multimedia online technologies as an exercise in translating materials into another medium, mostly for access and alternative to face-to-face or printed content delivery. This perception of technology use does not foster pedagogical change. It leads to

Table 1. Comparison of stated key competencies cited by teaching staff as important for students to acquire for their subject area to staff member's stated intended use of the online materials

Key competencies as stated by individual educators for students in their by individual	**Intended use of the online materials as stated educators teaching/subject area**
Educator A • Critical analysis • Ability to research • Standard academic writing skills	• *An adjunct to face-to-face teaching; students could access lectures if they could not come to lectures* • *It's the way things are going*
Educator B • Rhythmic perception • Rhythmic literacy • Programming skills	• *Wanted to have a more efficient way of doing things* • *Students access the materials (notes, exercises) before coming to lectures* • *To decrease degree of coordination because links are made obvious on the Web page* • *Access to materials off campus*
Educator C • Challenge their assumptions • Analyze the thinking, underlying practices • Connect theoretical material with their own life experiences • Think through how values can be incorporated into a real-life situation	• *A Web site that students could move around in rather than work linearly and that would get them thinking; to really engage them* • *Wanted to use class contact time for students to engage with each other on the basis of content they already encountered rather than using time for presenting content alone*
Educator D • Analytical skills • Mathematical skills • The case study approach is commonly used.	• *Resource that would be accessed in tutorials* • *To reduce but not replace lecture hours eventually*
Educator E • Develop problem-solving skills • Understand the material covered rather than just memorize it, and then apply what they have been taught to new situations • Become more creative in the tasks assigned	• *The key advantage to the students was greater accessibility and a more convenient way of delivering of course materials* • *Through supplementary activities such as reading, research, foresee what is going to be taught and contribute more to the class, rather than "being a clean slate" when material is presented*
Educator F • Analysis, synthesis, creativity • Develop an analytical way of thinking and problem analysis	• *Resource would have the same attributes as opening a book* • *Students have access to the content, but it really is only an add on*

counterproductive strategies that replicate more traditional methods with the new medium. The result is no impact or even negative impact on the learning environment. Rather, what is required is conceptualization of multimedia technology use in educational contexts as a process of transformation that acknowledges, and strives for, change in practice. In addressing this problem, it is useful to consider the idea of progressive technology adoption found in the literature.

Sandholtz, Ringstaff, and Dwyer (1997) suggested that supplemental use of multimedia technologies as was observed in this study should be viewed as the first stage of a continuum of change that culminates in a third stage of full integration and transformation of practice. The idea of progressive technology adoption is supported by others. For example, Goddard (2002) recognized five stages of progression: knowledge (awareness of technology existence); persuasion (technology as support for traditional productivity rather than as curriculum related); decision (acceptance or rejection of technology for curriculum use—acceptance leading to supplemental uses); implementation (recognition that technology can help achieve some curriculum goals); and confirmation (use of technology leads to redefinition of the learning environment—true integration leading to change).

It is proposed here that framing the educational use of multimedia technologies in terms of progressive levels of use and integration is valuable in that it forces conceptualization of effective technology integration as a process of "change" inherently leading to practice transformation rather than as simple skill acquisition required for translation of materials into a new medium.

Adopting the view that technology integration is a process leading to transformation and innovation directs attention also to the need to include elements of reflective practice in any educator development guidelines and frameworks. The term "reflective practice" is being used here to encompass the idea that educators consciously make judgments about their performances and success of strategies. The notion of evaluation (both formal and informal) is inherent in the idea of reflective practice. According to Ballantyne, Bain, and Packer (1999), lack of reflection leads to lack of awareness of the "appropriateness of…methods in bringing about high quality student learning" (p. 237), resulting in the perpetuation of traditional or ineffective teaching methods. The need for educators to reflect on their practices cannot be understated. Development of new strategies that appropriately integrate multimedia technologies into the curriculum will only take place, according to Tearle, Dillon, and Davis (1999), when the educator has "re-examined his or her approach to teaching and learning" (p. 10).

In the 2000 study conducted by Torrisi and Davis, another key finding was that among the concerns about the production process by educators, the principal concern was the lack of knowledge about the attributes and possibilities of the media and feelings of inadequacy in terms of how to exploit the potential of the media available. Consistent with other findings on professional development (Ellis, O'Reilly, & Debreceny, 1998), it appears that educators are primarily interested in learning the technical aspects of multimedia technologies only insofar as this knowledge is useful in informing pedagogical decisions and options. The implication of this observation is that teaching development efforts aimed at effective integration of multimedia technologies in educational contexts must teach educators how to use the technology within the context of "matching the needs and abilities of learners to curriculum goals" (Gonzales et al., 2002, p. 1).

The view upheld in this chapter is that using multimedia technologies within the curriculum context implies appropriate use of technologies. This view of appropriate technology use supports a mixed-mode approach to curriculum design. That is, the emphasis is on exploiting the attributes of various multimedia technologies and other strategy options in terms of their appropriateness

to content requirements, context, learner needs, and curriculum goals. Some guidelines and a development framework that encapsulate these views are discussed below.

GUIDELINES AND A DEVELOPMENT FRAMEWORK

In the discussion above, some key issues to be addressed in teacher development resources and approaches have been identified by drawing upon data from an earlier study (Torrisi & Davis, 2000). The author's perspective on addressing those issues was also alluded to. Drawing on issues identified in the preceding sections, this section presents the following:

1. A set of guidelines useful for guiding educator development activities
2. A planning framework that may be used to guide teacher development or by individual teachers in order to facilitate the effective integration of multimedia technologies in learning environments

A brief case study is also described in order to illustrate implementation of the notions presented.

Educator Development Activities: Five Key Guidelines

It has been established in the preceding sections that while multimedia technologies are seen as having the potential to reshape practice, the fact remains that implementation often results in little impact on the teaching space. The attributes of the multimedia technologies are not effectively exploited to maximize and create new learning opportunities. At the crux of this issue is the failure of educators to effectively integrate the multimedia technologies into the learning context. The following guidelines are suggested for guiding educator development toward the effective integration of multimedia technologies into learning environments.

- ***Guideline 1:*** **The goal of implementing multimedia technologies into learning spaces is to exploit the attributes of multimedia technologies in order to support deeper, more meaningful learner-centered learning. Realization of this goal necessarily transforms the teaching and learning space.** The knowledge-delivery view of multimedia technologies must be challenged, as it merely replicates teacher-centered models of knowledge transmission and has little value in reshaping practice. Constructivism is the guiding philosophy.
- ***Guideline 2:*** **Transformation is only achieved through integration of multimedia technologies into the learning space.** Integration implies that technology use is inextricably linked with the total curriculum as opposed to the superficial add-on approach that is the result of a view of translation.
- ***Guideline 3:*** **Integration and subsequent transformation is achieved via an ongoing evolutionary process through which educators' knowledge of multimedia technologies draws more closely toward inextricable linkages with curriculum goals and the educator's knowledge of pedagogy.**
- ***Guideline 4:*** **Equipping educators with knowledge about the potential of the multimedia technologies must occur within the context of the total curriculum needs** rather than in isolation of the academic's curriculum needs.
- ***Guideline 5:*** **Evolutionary process leading to transformation and integration of multimedia technologies is fueled by sustained reflection on practice.** Sustaining reflection on practice from the beginning of

> endeavors in online materials development through to completion stages, after which debriefing and further reflection feed back into a cycle of continuous evolution of thought and practice. Collaborative work and sharing of experiences and ideas with other educators is also of benefit here.

In addition to the above guidelines, two considerations as identified by Torrisi and Davis (2000) are important to recognize as contributing to effective professional development conducive to long-term transformation in practice.

First, it is important that professional development programs are not designed in isolation of the educators operating context. Traditional training workshops removed from the immediate teaching context of the educator fail to be effective. Programs must empathize with and address concerns that arise from educators' earlier attempts at innovation through technology. Ongoing support opportunities, both technical and pedagogical, must be inextricably linked with educators' everyday practice. If appropriate technology use is to be a reality, then professional development must do as Fatemi (1999) stated:

…more than simply show teachers where in a curriculum they can squeeze in some technology.... Instead, it helps them learn how to select digital content based on the needs and learning styles of their students, and infuse it into the curriculum rather than making it an end in itself. (p. 1)

Professional development programs will be most effective, as Bennet et al. (1999) stated, if educators are able to "connect the use of new technology to their own teaching experiences" (p. 212). The planning framework described below focuses on these ideas.

Second, in order for educators to be willing to use multimedia technologies in the classroom, it is necessary that they feel confident in their use from a technical perspective. Hence, professional development programs need to provide opportunities for developing basic computer competencies necessary for developing confidence in using technology as a normal part of teaching activities. Again, it is stressed that learning technical aspects must occur not in isolation of educators' teaching contexts, but rather in parallel with and integrated with pedagogical development. In this way, acquisition of technical knowledge is appropriate to the needs of the educators and is thus more likely to be relevant.

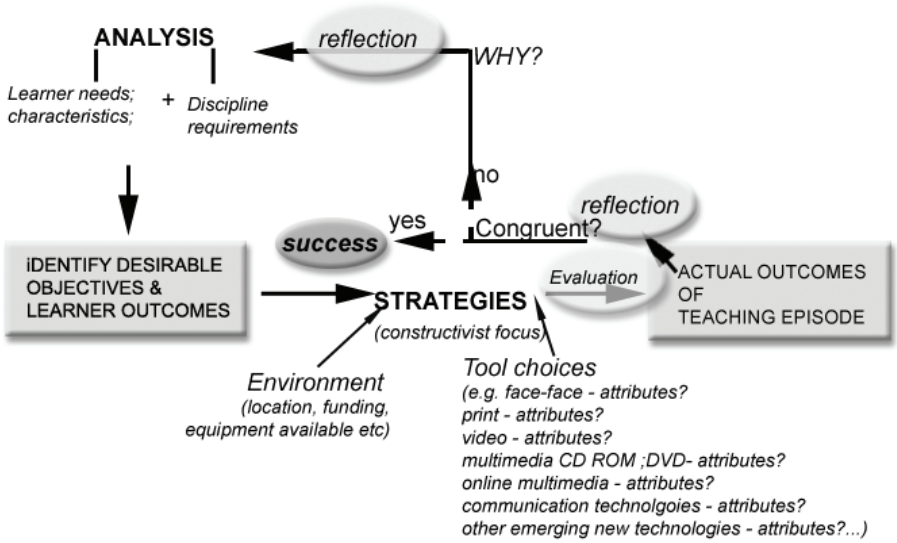
A Planning Framework

The five key guidelines above, together with issues discussed in this paper, can be embodied in a framework that provides a more concrete approach to curriculum planning conducive to the integrated use of multimedia technologies. The framework may be used to guide educator development (as has been done by the author) or may be useful as a guide for individual educators as they plan for multimedia technology use. The framework aims to highlight the use of multimedia technologies as part of the set of tools that is available for educators in executing teaching and learning strategies.

The framework is thus directed toward appropriate and judicious use of multimedia technologies. It also encourages educators to consider the attributes of them and then consider how to exploit those technologies for producing more meaningful and varied learning experiences; in so doing, allowing technology use to be an integral part of "knowledge spaces" which "allow users to explore as they wish" (Brown, 1997).

Consistent with learner-centered approaches, the process begins with an analysis of learner characteristics and of learner needs in relation to the content that is to be taught. In defining learner needs, the demands of the content must also be taken into account. As stated in Jamieson (1999), "The content of student learning (what is taught) logically precedes the method of teaching con-

Figure 1. Framework for appropriate integration of multimedia technologies into the learning environment. Environment attributes include human resources, financial resources, and other infrastructure and institutional limitations. Multimedia and other emerging multimedia technologies form part of the set of tool choices that the educator might choose on the basis that the attributes of the chosen tool(s) best fits with the learning context and desired outcomes.



tent…without content there is no teaching method" (p. 2). On the basis of this analysis, desired learner outcomes and objectives are identified.

In formulating teaching and learning strategies, the framework demands that the choice of tools be an informed choice based on integrated knowledge of strategy, learner needs, content requirements, environmental constraints (location, available equipment, funding, etc.), and tool attributes. Thus, the aim of technology integration more naturally precipitates from using the framework.

The fundamental view expressed by the model is that multimedia technologies and other emerging technologies are part of a tool set that, along with other available options (face-to-face teaching, print, etc.), are available choices for strategy implementation. Inherent in the presented framework is the philosophy that learning about multimedia technologies is an exercise in identifying the attributes of that technology and, at the same time, considering those attributes in terms of usefulness in the curriculum (Table 2). This approach addresses the problem of the "blanket approach" to multimedia technologies use that sometimes arises when the hype surrounding a new technology emphasizes the technology itself rather than learning as the primary concern. The proposed framework does not exclude the use of more traditional approaches or tools such as print, etc., if they are deemed appropriate to the learning situation.

It is also worth noting that the framework encourages evaluation of strategy outcomes and reflection on existing as well as new strategies. Consideration of the use of multimedia technologies occurs with the goal of modifying or replacing existing strategies that, upon reflection, are considered ineffective. This is considered an important characteristic of the framework for two reasons:

Table 2. Some common tool choices for teaching and learning and their main advantages and disadvantages

Tool	**Advantages**	**Disadvantages**
Textbook and other print materials	Portable, inexpensive, simple, "low-tech," easy to use, preorganized quantities of information, accessible without special equipment, can be inexpensive, most educators familiar with this medium and have production expertise	Become outdated, cannot update easily, static information presentation, no interaction possibilities
Video	Motivational, sound and images to convey information, readily available, easy to use, inexpensive	Linear information presentation, multiple copies for student access at home can be problematic/expensive, video production can be expensive and time consuming, requires VCR access and display mechanism
Face-to-face teaching	Can respond to needs of students dynamically, can be used to promote discussion, collaborative learning, enables clarification and analysis of information	No flexibility for students in terms of attendance, access limited to on campus
Multimedia CD ROM	Can convey information using video, audio, sound and text; once produced, inexpensive to replicate for student access; option for nonlinear information presentation, so learners are able to explore at their own pace, forming their own pathways; high interactive learning potential	Costly both in terms of time and money to produce; production requires a high level of technical expertise; software and content become outdated—cannot update easily without undergoing another development and production run
World Wide Web—Web sites and related Internet technologies	Increasingly supporting multimodal presentation—text, images, sound, video, and higher levels of interactive possibilities; access to up-to-date information; potential for collaborative learning with learners in multiple locations (e.g., chat, videoconferencing); potential for anytime, anyplace; highly motivational; updating of information relatively easy Wealth of up-to-date information available along with nonlinear nature, interactivity and multimodal presentation can support discovery orientated strategies	Requires costly technical infrastructure (networks, workstations, video conferencing facilities) Development of own online materials: complex requiring expertise; can be costly and time-consuming; involves a high level of commitment Updating Web materials can be difficult/frustrating if not technically competent to some degree Sophistication of Web materials available to students is limited by access factors such as bandwidth, modem capabilities Not all educators are familiar with/comfortable with the new media technologies—steep learning curve both in technical understanding and implementation strategies; lack of awareness of these issues is one of the greatest pitfalls in adopting multimedia technologies; **as technology capabilities increase, so do complexity, commitment required, and the potential of "things not working"**

1. It promotes the perspective that multimedia technologies are implemented with the primary goal of pedagogical change (thus helping to dispel the idea of a simple translation approach to technology adoption).
2. This encourages educators to draw on prior knowledge and experiences with prior teaching and make stronger connections between these experiences and the use of the technology (Bennet, Priest, & Macpherson, 1999). This is an important aspect of professional development efforts aimed at facilitating technology adoption for two reasons: perceptions of relevance are increased; and feelings of inadequacy that may be experienced by educators in dealing with new technology are minimized (Torrisi & Davis, 2000).

Reflective practice forms the cornerstone of the framework and is consistent with the notion of an evolutionary approach to technology integration. Execution of the strategy must be followed by a careful analysis of congruency of intended and actual outcomes. This analysis may involve formative and summative evaluation methods as well as personal reflection. The key question now becomes the following: Are the desired/anticipated outcomes congruent with actual outcomes? If they were, then the strategy is a success. Any discrepancies, however, need to be considered in the light of reflection of the process—Why did the discrepancies occur? In what ways might the strategy be changed or improved? Were the tool choices appropriate? The approach thus leads to a cycle of reflection followed by modified implementation followed again by reflection. Reflection on the process is not limited to assessment of whether outcomes were satisfactory, but rather encourages inspection of each stage of the planning process in order to identify shortcomings in either analysis or strategy. Aside from facilitating better technology integration into the curriculum, this approach may assist in overcoming some of the resistance to technology adoption, in that technology adoption becomes motivated by the need to improve practice.

The brief case study below illustrates an application of the framework and the beliefs expressed in this chapter.

A Case Study

The following case study is for a course in human services at tertiary level study. The case study summarized in Table 3 illustrates the key tenet of the framework described above that requires decision making regarding the tool choices for strategy implementation to be based on consideration of learner characteristics, desired learner outcomes, discipline requirements, and environmental considerations.

Resulting Subject Form

Given the adequate technical infrastructure, a subject Web site is used as the principal organizing medium for the subject and also as the primary means of preparation before engaging in face-to-face contact time. The Web site outlines weekly schedules and presents appropriate simulations and interactive exercises to introduce learners to course content and begin the process of self-reflection on beliefs and practices. The potential of multimedia-based Web sites to be used for more dynamic and engaging presentation of content prior to class time is exploited. Interactive case scenarios are presented via the Web site where they encourage students to explore their existing knowledge. This will "free-up" face-to-face class time for more valuable, deeper discussions rather than pure content presentation and initial reflection. Participation in discussion is important in helping students to analyze their own assumptions and in exposing them to the feelings and thoughts of others. The sharing of experiences, particularly after practical placements, is an important mechanism in this subject. Face-to-face contact is seen as an important tool for achieving

Table 3. Based on the planning framework described, this table illustrates how, for a tertiary-level human services course, decision making about tool choices for implementing strategies satisfies the constraints and demands of environment, learner characteristics, discipline requirements, and desired outcomes as well as addresses issues with previously used strategies. From the perspective of multimedia technology use, multimedia technologies are exploited in terms of attributes that will satisfy these demands and constraints. This facilitates appropriate and integrated technology use.

Tool choice (indicated by *) / **Issue/consideration**	Multimedia-based Web site	Web communications technologies	Print	video	Face-to-face
Environment: Good technical infrastructure (computer laboratories and Internet access) allowing for on-campus access outside working hours; the majority of students have computer and Internet access at home; regular on-campus contact time is also scheduled	*	*		*	*
Learner characteristics Mostly mature (nonschool learners) learners, with a high proportion of learners in full-time employment, with a cross section of abilities, backgrounds, and experiences; on-campus attendance is sometimes problematic	*	*	*		
Desired learner outcomes • "Challenge their own assumptions"	*	*			*
• "Analyze the thinking underlying practice"		*			*
• "Connect theoretical material with their own life experiences"	*	*			*

continued on following page

Table 3. continued

Tool choice (indicated by *) / **Issue/consideration**	Multimedia-based Web site	Web communications technologies	Print	video	Face-to-face
• "Think through how values can be incorporated into a real-life situation"	*			*	*
Discipline requirements: Off-campus practicum sessions— need for easily portable materials and off-campus access as well as communication with peers off-campus; thinking through and changing beliefs is a core goal of the subject	*	*	*	*	*
Issues with previously used strategy relying on face-to-face contact with print materials: Students failing to engage with reading materials prior to class, so much of valuable face-to-face contact time is used for delivery of content rather than active discussion Some students miss out on classes on occasion, because they are working or have other commitments	* *Web multimedia materials More interactive and stimulating— encourage more active involvement with content*	*			

learner outcomes that focus on analysis of beliefs and practices.

The Web site also enables students to have around-the-clock access to class materials. Chat rooms are useful in enabling students to support each other and collaboratively solve problems, especially when away from campus on practicum. Support is also available from tutors at certain times of the week, during which students can log in while off campus. This is an important mechanism for helping students to connect theoretical knowledge with the experiences they are undergoing at the time. E-mail is also useful in encouraging and maintaining student–faculty contact during practical and other off-campus times.

Print materials will still be used but primarily as an easily portable reference source, especially while on practicum, rather than as a primary source of information prior to attendance at face-to-face class times.

FUTURE TRENDS

The range and nature of multimedia technologies are constantly changing. In an educational context, the challenge is to develop approaches to planning that can be used to facilitate integration of existing and future technologies. By conceptualizing the role of technologies in the learning context as a component of a set of tools, it is intended that the framework and ideas presented in this chapter be one step in the direction of "generic" planning approaches that will provide guidance to educators and those involved in their professional development in both current and future technological environments. Planning approaches need to be contextually framed so that the key focus is to exploit the attributes with the aim of providing deeper, more meaningful learning experiences that will equip students with the lifelong learning skills demanded in the present and the future.

Evidently, with the dynamic nature of multimedia technologies, there will always remain a need for ongoing professional development that will present opportunities for educators to investigate the attributes of multimedia technologies as they emerge in terms of usefulness for their particular teaching contexts. From this perspective, ongoing research focusing on which technologies are being used in what contexts and what results are being obtained becomes important. Such research will highlight the attributes of the technology that are worth exploiting and for what purposes and could result in models of implementation so that educators could draw upon one another's experiences.

CONCLUSIONS

The framework and ideas presented in this chapter precipitated from concern about the ineffective and often inappropriate use of multimedia and associated technologies in learning contexts.

The fundamental belief expressed in this chapter is that effective use of multimedia and other technologies in teaching and learning environments occurs when multimedia technologies are integrated fully and appropriately into the curriculum. The primary goal for integrating multimedia and associated technologies into the curriculum is to provide for a learning environment that espouses more meaningful and deeper learning.

It is advocated throughout this chapter that multimedia and associated technologies are considered as part of a tool set available for strategy implementation. If, how, and when to integrate technologies are decided by taking into account constraints and conditions imposed by the environment, learner characteristics, desired learning outcomes, and the nature of the content, and by reflecting on success or otherwise of previously used teaching practices. Also highlighted is the important role of reflective practice. Another key

theme is the need to foster the view that technology integration is an evolutionary, transformative process rather than an exercise in translation of strategies to another medium. The five guidelines for multimedia technology use and the planning framework presented in this chapter incorporate these views.

The dynamic, rapidly evolving technological environment characteristic of the present and the future represents ongoing challenges for educators striving to make use of these new tools to the best advantage for a more effective learning environment and more meaningful learning outcomes. Despite the dynamic nature of technical environments, it is the author's belief that there is at least one constant premise upon which educator development aimed at multimedia technology integration efforts can develop—that is that change in practice is inextricably linked with successful integration of multimedia technologies in teaching and learning contexts. Nurturing the acceptance of this premise needs to be a matter of priority in current and future educator development efforts in the area of educational application of technologies.

REFERENCES

Ballantyne, R., Bain, J. D., & Packer, J. (1999). Researching university teaching in Australia: Themes and issues in academics' reflections. *Studies in Higher Education, 24*(2), 237–257.

Bennet, S., Priest, A., & Macpherson, C. (1999). Learning about online learning: An approach to staff development for university teachers. *Australian Journal of Educational Technology, 15*(3), 207–221.

Brown, T. (1997). *Multimedia in education—Conclusions.* Retrieved September 27, 1999 from the World Wide Web: http://129.180.87.4/Units/CurricSt/CSIT513/573/573_12.html

Burns, M. (2002). From black and white to color: Technology, professional development and changing practice. *T.H.E. Journal, 29*(11), 36–42.

Charp, S. (2000). Technology integration. *T.H.E. Journal, 29*(11), 8–10.

Collis, B. (1996). *Pedagogy.* Retrieved September 10, 2002 from the World Wide Web: http://www2.openweb.net.au/TT96University/BC.html

Conlon, T., & Simpson, M. (2003). Silicon Valley versus Silicon Glen: The impact of computers upon teaching and learning: A comparative study. *British Journal of Educational Technology, 34*(2), 137–150.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920.* New York: Teachers College Press.

Ellis, A., O'Reilly, M., & Debreceny, R. (1998). *Staff development responses to the demand for online teaching and learning.* Paper presented at ASCILITE '98 conference, Wollongong. Retrieved March 20, 2003 from the World Wide Web: http://www.ascilite.org.au/conferences/wollongong98/ascpapers98.html

Fatemi, E. (1999). *Building the digital curriculum. Education Week on the Web.* Retrieved July 16, 2001 from the World Wide Web: http://www.edweek.org/sreports/tc99/articles/summary.htm

Goddard, M. (2002). What do we do with these computers? Reflections on technology in the classroom. *Journal of Research on Technology in Education, 35*(1), 19–26.

Gonzales, C. L. P., Hupert, N., & Martin, W. (2002). The Regional Educational Technology Assistance Program: Its effects on teaching practices. *Journal of Research on Technology in Education, 35*(1), 1–18.

Hammond, M. (1994). Measuring the impact of IT on learning. *Journal of Computer Assisted Learning, 10*, 251–260.

Jackson, B., & Anagnostopoulou, K. (2000). *Making the right connections: Improving quality in online learning. Teaching and Learning Online: New pedagogies for new technologies. International Centre for Learner Managed Learning, Middlesex University.* Retrieved April 15, 2003 from the World Wide Web: http://webfeedback.mdx.ac.uk/_lmlseminar/_private/_abstract14/finland.htm

Jamieson, P. (1999). *Improving teaching by telecommunications media: Emphasising pedagogy rather than technology.* Paper presented at the Ed-Media 1999 World conference on Educational multimedia, hypermedia and telecommunications, Charlottesville.

Lefoe, G. (1998). *Creating Constructivist learning environments on the Web: The challenge of higher education.* Paper presented at ASCILITE '98 conference, Wollongong. Retrieved March 20, 2003 from the World Wide Web: http://www.ascilite.org.au/conferences/wollongong98/ascpapers98.html

Nichol, J., & Watson, K. (2003). Editorial: Rhetoric and reality—The present and future of ICT in education. *British Journal of Educational Technology, 34*(2), 131–136.

Oliver, R. (1999). *Teaching and learning with technology: Learning from experience. In On the Edge Leading the learning revolution.* Paper presented at the Proceedings of the Australian Curriculum Assessment and Certification Authorities Conference, Perth.

Pierson, M. E. (2001). Technology integration practice as a function of pedagogical expertise. *Journal of Research on Computing in Education, Summer,* 413–430.

Rakes, G. C., & Casey, H. B. (2002). *An analysis of teacher concerns toward instructional technology.* International Journal of Educational Technology. *3*(1). Retrieved March 30, 2003 from the World Wide Web: http://www.outreach.uiuc.edu/ijet/v3n1/rakes/index.html

Relan, A., & Gillani, B. (1997). Web-based instruction and the traditional classroom: Similarities and differences. In B. H. Khan (Ed.), *Web-based instruction.* New Jersey: Educational Technology Publications.

Richards, C., & Nason, R. (1999). *Prerequisite principles for integrating (not just tacking-on) multimedia technologies in the curricula of tertiary education large classes.* Paper presented at the ASCILITE '99 Conference. Brisbane. Retrieved March 30, 2003 from the World Wide Web: http://www.ascilite.org.au/conferences/brisbane99/papers/papers.htm

Sandholtz, J., Ringstaff, C., & Dwyer, D. (1997). *Teaching with technology.* New York: Teachers College Press.

Selwyn, N., & Gorard, S. (2003). Reality bytes: Examining the rhetoric of widening educational participation via ICT. *British Journal of Educational Technology, 34*(2), 169–181.

Strommen, D. (1999). C*onstructivism, technology, and the future of classroom learning.* Retrieved April 15, 2003 from the World Wide Web: http://www.ilt.columbia.edu/ilt/papers/construct.html

Tearle, P., Dillon, P., & Davis, N. (1999). Use of information technology by English university teachers. Developments and trends at the time of the National Inquiry into Higher Education. *Journal of Further and Higher Education, 23*(1), 5–15.

Torrisi, G., & Davis, G. (2000). Online learning as a catalyst for reshaping practice—The experiences of some academics developing online materials. *International Journal of Academic Development, 5*(2), 166–176.

Torrisi-Steele, G. (2001). Appropriate use of multimedia technologies in tertiary learning environments. *Staff and Educational Development International*, *5*(2), 167–176.

This work was previously published in Interactive Multimedia in Education Training, edited by S. Mishra and R. C. Sharma, pp. 25-46, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.7
Planning Effective Multimedia Instruction

Chien Yu
Mississippi State University, USA

Angela Williams
Mississippi State University, USA

Chun Fu Lin
Mississippi State University, USA

Wei-Chieh Yu
Mississippi State University, USA

ABSTRACT

Multimedia benefits students learning in many different ways. There are so many things that students can do and learn because of the variety of instructional media that is available for their use. The use of instructional multimedia increases an instructor's ability to propose and execute teaching strategies that come with a multiplicity of learning styles. Therefore, there are a myriad of reasons why teachers use these resources both as a teaching tool and a teaching resource. Several strategies can be implemented so that teachers have opportunities to become skillful in attaining technological fluency. This chapter reviews the trends and issues of today's multimedia education, and attempts to provide strategies and guidelines for planning multimedia instruction. The effective use of pedagogical design principles with appropriate multimedia can permit greater individualization, which in turn fosters improved learning, greater learner satisfaction, and higher retention rates.

INTRODUCTION

Technological capabilities are growing in today's world by leaps and bounds. Over the past few decades, there have been remarkable advances in

computer and interactive media technology. As a result, there has been a tremendous increase of investment in school technology and media use. Teachers are also being asked to learn the skills and techniques required to use computers and instructional media in classrooms.

The diverse characteristics of different multimedia and the capabilities that they provide for learning have direct implications on the design of multimedia strategies and materials (Fahy, 2005). Goldman and Torrisi-Steele (2005) state that "the essential value of interactive multimedia technologies is that they can be used effectively to empower students to take a more pro-active role in acquiring, analyzing, and synthesizing information" (p. 191). Although today's technologies make possible the use of multimedia by helping to move learning beyond a primarily text-based and linear arena into the cyclical world of sights, sounds, creativity, and interactivity, the challenge is whether the essence of multimedia can be integrated into an essential discipline (Gonzalez, Cranitch, & Jo, 2000). If some pedagogical design principles and appropriate media are used effectively, multimedia can permit greater individualization, in turn fostering improved learning, learner satisfaction, and retention rates (Fahy, 2005).

This chapter discusses the definition of "multimedia," including the trends and issues of today's teaching and learning. The goal is to review the benefit of using multimedia instruction, and to attempt to provide some strategies and guidelines for planning multimedia instruction. By outlining some fundamental issues and considerations affecting implementation of multimedia, the chapter discusses some challenges and impacts of multimedia instruction when teachers are ready to move from simpler to more complex combinations of media for teaching. Additional examples drawn from literature are also included to discuss the use of multimedia in education and the strategies of planning effective instruction.

TRENDS AND ISSUES

The rise in the usage of technology is bringing about rapid change in the educational environment. In keeping with this changing environment, teachers need to discover ways to broaden their range of teaching methods so that they can produce more effective learners. Emerging trends including individualized learning, cooperative learning, collaboration learning, learner center approach, and assessment portfolio have been playing an important role in education. Research indicates the importance of increased technology integration in the classroom. When using interactive technology, students not only learn more quickly and enjoyably, but also learn the much needed life skill of learning how to learn (Vogel & Klassen, 2001). However, many educators today are facing the issue of integrating technology into their instruction (Wang & Speaker, 2002).

Technology continues to change dramatically. Although it may be recognized by educators that multimedia technologies have the potential to offer new and improved learning opportunities, many educators fail to realize this potential (Torrisi-Steele, 2005). Similarly, Kaufman (2002) also summarizes that most teachers have been taking advantage of technology's mass storage capacities, but they have not exploited its greater potential to motivate knowledge construction and facilitate problem solving. As a result, a number of educators using multimedia technologies in their learning environments are mainly limiting its use to a tool for data access, communications, and administration (Conlon & Simpson, 2003) rather than a tool for integrating curriculum (Torrisi-Steele, 2005). This lack of true integration results in minimal change in both pedagogical strategies and learning environment (Tearle, Dillon, & Davis, 1999).

In addition, failure to implement effective technology integration could be associated with teachers' technology skills and attitude as well. As Speaker (2004) concludes, student use and perception of a multimedia educational

experience is highly dependent on the attitude of the teachers and their ability to provide useful contextual information in a format that meets the criteria of relevancy and interactivity in a student-centered approach. However, research (Torrisi-Steele, 2005) shows that teachers are generally unprepared for the technology changes, and often lack the skills as well as technical and pedagogical knowledge to effectively implement those technologies in their learning environments. As a result, students' classroom practice may not meet student expectation especially in the area of integration and use of multimedia (Speaker, 2004) because today's students are often far more skilled at using digital media than most of their teachers. Torrisi-Steele states (2005), "The effective integration of multimedia in the curriculum depends not on the technology itself but rather on educators' knowledge, assumptions, and perceptions" (p. 26). Therefore, teachers should develop a desire to integrate technology into their classroom, and also need careful plans for multimedia instructions.

WHAT IS MULTIMEDIA?

The term "multimedia" means different things to different people; however, definitions of multimedia tend to agree in substance (Fahy, 2005). Fahy (2005) states that "the term "multimedia" refers to the provision of various audio and video elements in teaching and training materials" (p. 3). When Ivers and Barron (2002) define multimedia as "the use of several media to present information" (p. 2), Mayer (2001) views multimedia as "the presentation of material using both words and pictures" (p. 2). To Peck (as cited in Speaker, 2004) multimedia is "as a computer controlled combination of two or more media types, to effectively create a sequence of events that will communicate an idea visually with both sound and visual support" (p. 242). Not only do Roblyer and Schwier (2003) define multimedia as "a computer system or computer system product that incorporates text, sounds, pictures/graphics, and/or audio" (p. 329), but they also imply its purpose as one of "communicating information."

Due to the growing delivery of media by the computer and the merging of increasingly powerful computer-based authoring tools with Internet connectivity, it seems that the term "multimedia" is now firmly associated with computer-based delivery. Although the term has not always been associated with computers (Roblyer & Schwier, 2003), Gonzalez et al. (2000) write that "multimedia cannot be experienced without the technology because it is the technology that creates the experience" (¶ 9). Therefore, Gayeski (1993) defines computer-based multimedia as "a class of computer-driven interactive communications systems which create, store, transmit, and retrieve textual, graphical and auditory networks of information" (p. 4). In other words, computer-based multimedia involves the computer presentation of multiple media formats (e.g., text, pictures, sounds, video, etc.) to convey information in a linear or nonlinear format (Ivers & Barron, 2002).

With an increased availability of digital information options, today's commonly accepted definition views multimedia as a combination of different media (i.e., text, pictures, sounds, video, animations, etc.) used to present multimodal information, in conjunction with computer technology. However, the majority of definitions, in general, only take into consideration the basic instructional delivery system. These definitions may be less meaningful, particularly in education, if multimedia cannot be incorporated with the way people learn and work. Because multimedia enables learners to engage the greatest number of the senses, multimedia can be a very powerful pedagogical tool if teachers can focus on the foundation for learning. By integrating multimedia to teaching and learning, students can work with increased information in more creative ways than ever before to make the knowledge more applicable and retainable.

MULTIMEDIA LEARNING

The capability of multimedia provides teachers with an array of learning pathways to offer students. According to Biggs (1999), learning is a way of interacting with the world. In other words, learning takes place through the active construction of knowledge by the interaction between information received through different channels and existing knowledge stored in the learner's long-term memory (Christie & Collyer, 2005). Wild (1996) describes what is successful in learning is to describe successful interactions between learner, context, and instruction. The potential of multimedia is to foster the level of interactivity as a form of learning, and to offer many possibilities for enriching the knowledge or information for learners. As Zhang (2005) concludes, multimedia instruction along with high levels of interaction can help maximize learners' ability to retain information and learner engagement.

Benefit of Multimedia Instruction

There are many benefits of interactive multimedia instruction. The first is compliance with the requirements of No Child Left Behind Act of 2001 (NCLB). The challenges of NCLB no longer allow schools the luxury of viewing technology as an add-on function at the periphery of instruction, a curriculum enhancement, or as an occasional frill that makes learning more fun; instead, schools and particularly teachers must adopt and embrace technology so that it restructures the way learning is managed and administrated (Tetreault, 2005). The increased control of pace, sequence, difficulty, content, and/or style of the instructional media presentations allows learning to be customized to each student's distinctive needs. Teachers can thus ensure at least some progress of the student. The students, on the other hand, can better comprehend the level of skill and knowledge that is required by the work, and also progress at their own pace, allowing for greater achievement by those who are mentally ready to move ahead. This is an important step based on the NCLB Act, which calls for the Secretary of Education to "conduct a rigorous, independent, long-term evaluation of the impact of educational technology on student achievement using scientifically based research methods and control conditions" (as cited in Means & Haertel, 2004, p. 9). This call for action is to evaluate the effects of the investment of bringing instructional media into the classroom. If the Education Secretary finds positive results in the use of instructional media in the classroom, then it shows that the national investment has indeed been a worthwhile venture.

Another benefit of media-based instruction is the reduction of the time required to reach instructional objectives. If teachers do not have to spend instructional time repeating material that most of the students already know, then the learning of new material could occur more quickly. Since media-based instruction also keeps students interested and more involved in educational material, this allows for wiser use of both their time as well as the teacher's time.

Multimedia instruction also provides a stimulus for learning by increasing social interactions and cooperation. Hoyles (1994) reports that there is a relationship between students involved in multimedia-based environment where the collaboration was seen to lead to higher-order thinking, hypothesis formation, and reflection. A Mevarech and Light (1992) report on a review of studies, which investigated the potential of multimedia-based learning environment to enhance group work provides convincing evidence of the value of group work and collaboration, and its positive impact on productive learner dialogue, interchange of ideas, and negotiation of solutions.

With one-to-one interaction being regarded as an educational ideal, efforts to realize this through information technology have included intelligent tutoring systems and telementoring. These intelligent tutoring systems, like their human counterparts, are fully expected to respond

flexibly to student inputs so as to optimize progress toward a learning objective. It involves one-to-one interchanges between tutor and student, typically relying on e-mail exchanges between a student and experts in that field. The program's success is extremely dependent on the match between the mentor and the student (Kovalchick & Dawson, 2004).

New knowledge media such as video productions, simulations, and microworlds extend the range of experiences and concepts that can be brought into school. For example, field trips are regarded as classic ways to explore worlds not easily represented through school-based instructional materials. With today's virtual reality, students can experience an interactive "field trip" without leaving their classroom. In addition, providing multimedia instruction has been found to have a positive impact on students' perceptions of teachers. For example, teachers who use presentation visuals are judged to be better prepared, more concise, professional, clearer, credible, and interesting (Vogel & Klassen, 2001). Also, multimedia technology can enable teachers to work together to share material and complement each other's expertise, thereby adding value to education (Alavi, Yoo, & Vogel, 1997). The use of multimedia also links students and classes together as well as enhances the ability of government and business experts to participate in education.

Use of Multimedia Instruction

The strength of multimedia instruction lies in its ability to adapt to students' individual differences and capabilities of controlling the learning path. With the new trend in education that emphasizes the importance of learning *with* technology instead of learning *from* technology (Jonassen, Howland, Moore, & Marra, 2003), the accessibility of multimedia technologies has presented an array of choices to instructors these days. For example, there are many ways that computer use can and does benefit students in their learning. The first major instructional use of computers and software programs was simply for drill and practice. This was seen as a beneficial use for students who need extra or special practice when the teacher just did not have enough time to perform this task.

Another use is computer-assisted instruction. This requires more advanced instructional programming than drill and practice programs. This type of software provides concepts and content in a straightforward approach. As Chipman (2003) indicates, "The computer offers management of the student's study efforts through pacing and interspersed questions" (p. 39). These programs can vary from simple to more sophisticated where the computer may respond differentially to student responses with preprogrammed responses of its own.

An additional use, which is also quite popular, is the use of simulations. These are safe environments and are easily accessible to students. "Simulations make it possible to experience an approximation of phenomena that otherwise might merely be talked about" (Chipman, 2003, p. 40). Simulations open new possibilities in teaching and permit students to practice events or phenomena that might be too hazardous or too intricate to duplicate, but to be effective, they must be integrated with a larger curricular context. For example, in a science class, if students are learning the body parts of frogs, students can use a computer program to simulate a frog dissection, instead of cutting up an actual frog. This simulation may not give students the hand-on experience; however, it is much cheaper and it can be used over and over. It is also more flexible because the frog can be reassembled by the program, and the simulation can avoid sacrificing a real frog.

Another popular use today is streaming video. It revolutionizes the way visual content is used in the classroom for clear demonstrations of proficient performance. It also provides more learning potential that the use of VCR's, and empowers teachers and students to study a broader range of topics with higher retention and comprehension,

which is a key to maximizing student achievement (Holland, 2005). With streaming video, students can now watch a lecture, tutorial, or presentation online via normal modem, and at the same time review the corresponding documents or materials in a highly compressed, Web browser accessible format.

Other uses that are applicable from elementary school level and up include drill and practice software, problem-solving software, Web-based software, data analysis and reporting software, and demonstration presentation software (Bitter & Pierson, 2003). These are all tools that can and should be very valuable assets for many grade levels and subject areas. For example, presentation software such as PowerPoint allows teachers to place class materials together in a dynamic and meaningful manner. The ability to integrate sound, picture, and video into PowerPoint slides offers tremendous potential to captivate the child's attention and give the child a sense of control over the learning process that makes it more palatable (Yu & Smith, in press). Interactive applications such as the World Wide Web can be used as a main information resource for students to access any time and any where, or for the storage of instruction materials including class notes, presentation visuals, and video recordings. In addition, data analysis and reporting software such as Excel provides all the functions to manipulate numbers, takes numeric and alphanumeric format data as input to a matrix, performs matrix calculations, and produces a computed spreadsheet as the main output. For instance, students in math and finance related classes might be able to use Excel to perform predefined calculations automatically, provide statistical functions, and display charts. In some ways, it is easier to use computers over calculators since numbers are visible all of the time on the screen. Students can move gradually from using them as a super-calculator to using them for modeling or simulation. In other words, they can move from using them to ask *What is* questions to *What if* questions. Thus, they also can be used as decision-making aids.

MULTIMEDIA FOR EFFECTIVE INSTRUCTION

The term effective instruction as it is used today does not mean the same thing that it did just a few years ago. It once meant simply teaching the basics—reading, writing, and arithmetic. Now it means much more. It means not only teaching the basics, but also teaching concepts and discoveries using technological media. There are so many things that students can do and learn because of the variety of instructional media that is available to them. There are a myriad of reasons why teachers must use these resources both as a teaching tool as well as a teaching resource. One reason is because of the ever-increasing amounts of interactive instructional media resources that have become available all over the world, even in our schools. Another reason is because students today are far more technology savvy than most teachers are. Additionally, most states have added technology proficiency to their teacher licensing requirements. Because of this, policy makers, teacher educators, teachers, and school administrators must do something to ensure that all teachers are proficient in the use of instructional media as educational tools (Zhao, 2003). Teachers are responsible for creating a learning environment that emphasizes the impact of content and increases student learning (Vogel & Klassen, 2001). With the assistance of multimedia technology, demands from the teacher in a technological learning environment are increasing also. Thus, the importance of planning effective multimedia instruction is continuously growing for individual educators in primary, secondary, and other training environments.

PLANNING FOR MULTIMEDIA INSTRUCTION

The value of integrating multimedia technology into classrooms at all levels has been discussed

in many research studies in different disciplines (Agarwal & Day, 1998; Stone, 1999). Gonzalez et al. (2000) indicate that "multimedia is more than a collection of sound, images, video and animations. It is a vital, dynamic field offering new challenges, interesting problems, exciting results, and imaginative applications" (¶ 5). They point out the challenge--whether the unique essence of multimedia can be distilled into an essential discipline (Gonzalez et al., 2000). Successful technology integration involves careful evaluation of the curriculum and learning goals (Ertmer, 1999). Therefore, the following section seeks to provide some strategies and guidelines for planning multimedia instruction. The pedagogical aspects of effective instruction are the main focus in this discussion.

State Technology Standards

The first logical and practical step for multimedia instruction is to look at educational technology standards for teachers at the state level. Most states should have these standards in place as part of their blueprints or guidelines on what should be minimally taught at each grade level. These objectives provide a starting point on effectively teaching the use of technology and media to students. However, teachers should also realize the need to incorporate these strategies into all subject matter and not treat it as a separate entity. They should also remember that these are minimal goals and that they can and should go further in their endeavor to use a variety of media to benefit them in teaching and students in learning.

Design and Development Principles

The quality of the learning experience depends considerably on the design and presentation of instructional materials (Sanders & Morrison-Shetlar, 2001). A number of studies provide principles that can guide teachers in the design and development of multimedia instruction. Mayer's (2001) work is one such example. He reveals that successful learning requires students to perform five actions, with direct implications on the design of effective multimedia instruction:

- Select relevant words from the presented text or narration.
- Select relevant images from the presented illustrations.
- Organize the selected words into a coherent verbal representation.
- Organize selected images into a coherent visual representation.
- Integrate the visual and verbal representations with prior knowledge (p. 53).

In order to guide the design of multimedia instruction, Mayer (2001) articulates seven useful principles that can help students to achieve greater retention.

1. **Multimedia principle:** Students learn better from words and pictures than from words alone.
2. **Spatial contiguity principle:** Students learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.
3. **Temporal contiguity principle:** Students learn better when corresponding words and pictures are presented simultaneously rather than successively.
4. **Coherence principle:** Students learn better when extraneous words, pictures, and sounds are excluded rather than included.
5. **Modality principle:** Students learn better from animation and narration than from animation and on-screen text.
6. **Redundancy principle:** Students learn better from animation and narration than from animation, narration, and on-screen text.
7. **Individual differences principle:** Design effects are stronger for low-knowledge learners than for high-knowledge learners

> and for high-spatial learners rather than for low-spatial learners (p. 184).

Learners are the core in the realm of teaching and learning. Since successful interaction design that engages learners in exploring knowledge and experiences is the result of careful analysis of the learner and of the learning outcomes (Goldman & Torris-Steele, 2005), Fardouly (as cited in Goldman et al., 2005) provides some questions to guide successful interaction design while constructing multimedia interactivities:

- Who are the learners? What do they need or want to learn? In what environments will the learning be applied, and what do they already know?
- What is the teacher trying to achieve with the instruction? Clearly define goals and objectives and relevant content.
- What skills, attitudes and knowledge are you trying to develop?
- How will content be structured?
- What strategies might be used?

These examples of design principles may enhance learners' learning experience and maximize the potential of multimedia technologies.

Educational Theories-Constructivism

Leidner and Jarvenpaa (1995) categorize learning models into several categories: *objectivism, constructivism, collaborative learning, cognitive information processing,* and *socioculturism*. Among them, the leading theory of today's learning is constructivism, which is the idea that learning actually occurs when learners actively try to understand material that is presented to them. They engage in constructivist learning by deeply and actively processing the material that is to be learned in an attempt to understand it. This process can also be called knowledge construction because learners create their own knowledge, apply, and coordinate it to their own cognitive processes while learning. This learning is also traditionally known as meaningful learning or learning by understanding (Mayer, 2003).

The constructivist view of teaching and learning is a commonly accepted framework for developing appropriate strategies for designing multimedia learning environments in ways, which will promote student-centered learning environments (Goldman & Torrisi-Steele, 2002). According to Savery and Duffy (1996), effective instructional design of multimedia interactivities may be based on eight constructivist principles. They are:

- Anchor all learning activities to a larger task or problem
- Support the learner in developing ownership for the overall problem or task
- Design an authentic task
- Design the task and learning environment to reflect the complexity of the environment that students should be able to function in at the end of learning
- Give the learner ownership of the process used to develop a solution
- Design the learning environment to support and challenge the learner's thinking
- Encourage testing ideas against alternative views and alternative contexts
- Provide opportunity for, and support reflection on, both the content learned, and the learning process itself (p. 3)

The concept of constructivist learning does indeed have important implications for the use of a variety of interactive instructional multimedia. It is aimed at fostering and guiding learning and activating cognitive processing that leads to understanding. "Under this conception of learning, instructional technology should serve as a cognitive guide to help learners on authentic academic tasks-such as comprehending a text, solving challenging

mathematics problem, or conducting a scientific experiment." (Mayer, 2003, p. 128)

Learning Styles

Learning styles are also a well-known area impact on today's technology-based instruction. As Kovalchick et al. (2004) state, "Learning styles are the diverse ways in which people take in, process, and understand information" (p. 418).

It is of importance to be aware of different learning styles possessed by the learners. Litchfield (1993) indicates, "…matching learning style with design of instruction was important for both achievement and positive attitudes" (p. 5). As educators, we need to know that one size does not fit all. One learner might learn best in a cooperative learning environment while the other may achieve similar learning outcome through self-study (Marlow, 2003). This does not at all mean that the teacher has to have classroom materials tailor-made just to satisfy one or two persons' special needs. The educator should, however, be sensible and flexible to different individuals, and include a variety of methods to account for different learning styles. The use of interactive instructional media in education increases an instructor's ability to propose and execute teaching strategies that cater to a variety of learning styles. The use of multimedia facilitates learning by providing a unique opportunity to notice and adjust to the differences in every person. The use of computer-mediated communications tools like e-mail, discussion boards, and virtual chat provide many opportunities for interaction collaboration and discussion inside as well as outside of the classroom (Kovalchick et al., 2004). Teachers can use interactive media to develop and deliver instruction to a variety of learners with diverse learning styles. E-learning, synchronous or asynchronous, that is conducted over the Internet, intranet, extranet, or other Internet-based technology is another growing trend in schools today that takes learning styles into account (Abram, 2005).

Using a variety of educational media also provides a stage upon which numerous instructional approaches can be developed. This fact relates back to Howard Gardner's theory of multiple intelligences. Gardner believed that each person has a different intellectual composition made up of verbal-linguistic (speaking, writing, and reading), mathematical-logical (reasoning skills), musical, visual-spatial, bodily kinesthetic, interpersonal, intrapersonal, naturalistic, and existential skills. People possess all the intelligences in varying

Figure 1. Examples of an interactive event at each functional level of interaction

	Confirmation	*Pacing*	*Navigation*	*Inquiry*	*Elaboration*
Reactive	Learner matches answer given by system	Learner turns page when prompted	Learner selects choice from a menu	Learner uses "help" menu	Learner reviews a concept map
Proactive	Learner requests test when offered	Learner requests an abbreviated version of instruction	Learner defines unique path through instruction	Learner searches text using keywords	Learner generates a concept map of the instruction
Mutual	System adapts to progress of learner and learner may challenge assessment	System adapts speed of presentation to the speed of the learner	System advises learner about patterns of choices being made during instruction	System suggests productive questions for the learner to ask given previous choices	System constructs an example based on learner input, and revises it as learner adds information

amounts and they may employ each one separately or jointly, depending on the learning situation. The variety of interactive media that we are provided with today (such as power point, e-learning, streaming video, the WWW, etc.) can easily be used to correspond with each individual's learning style and multiple intelligence area.

Meaningful Content and Interaction

Research (Esquivel, 1995) has shown that students enjoy the instructions related to real life events. Multimedia learning can support instruction that is contextually relevant, interactive, and meets the needs of the individual learner (Speaker, 2004). Multimedia can be made challenging and varying in degree of difficulty, yet the content should be relevant to student's study (Marlow, 2003). Adding adequate numbers of audios, visuals, and texts will aid successful knowledge construction if the instructional designer has taken the curriculum and learner's ability level into serious consideration.

The different multimedia instructions can emphasize different types of interaction. Schwier (1993) categorizes interaction as *reactive, proactive,* or *mutual* depending upon the level of engagement experienced by the learner. Hannafin (1989) identifies five functions interaction can serve in independent learning materials. They are *confirmation, pacing, inquiry, navigation,* and *elaboration.* Each function is expressed differently during instruction, depending on the level of interaction, and Schwier (1993, p. 168) provides one example of interaction obtained at each functional level of the taxonomy (see Figure 1).

Multimedia instruction needs to provide meaningful content for student learning. Teachers need to provide content, which makes sense to the student. Effective learning relies on meaningful interactions between content and learners. Meaningful interactions require the learner to access that meaningful knowledge in order to relate it to new information (Zazelenchuk, 1997). As a result, instructors can help learners access the appropriate knowledge and integrate it with the new information being presented by designing meaningful content and providing the right interactions.

Learner Control

The perception of learner control is another important element. Doherty (1998) refers to learner control as the level of self-determination that the learner has in making decisions about his or her learning, and more specifically defines it as "the degree to which individuals control the path, pace, and/or contingencies of instruction" (¶ 3). Lepper (1985) indicates that learner control may increase feelings of competence, self-determination, and intrinsic interest. Since multimedia can help present information in a nonlinear or random access format, the students can not only select what information to access and how to sequence the information in a manner that is meaningful to them (Lawless & Brown, 1997), but also make decisions about their own pacing and sequencing and follow through the interactive material during studying sessions (Sims & Hedberg, 1995).

Kinzie (1990) indicates that "exercising control over one's learning can be in itself a valuable educational experience" (p. 6). Lawless & Brown (1997) emphasize that "learners within a multimedia environment must not only understand the information presented, but must also be able to identify what information will further enhance understanding and how to access this information" (p. 121). Literature (Ross & Morrison, 1988) reveals that allowing learners to select text density and problem themes can have a positive effect on both performance and the perception of learner control. In addition, allowing the users free access to information may meet the needs of the learner and also positively impact attitudes about using the medium, or allowing the students to follow a specified path of information, choosing to revisit the information or to proceed onto the next

step may be more beneficial (Lawless & Brown, 1997).

Lawless and Brown (1997) indicate that learner control can positively influence effectiveness and efficiency of learning. They summarize five basic author imposed control levels, including *browsing, searching, connecting, collecting,* and *generating.* They also indicate that these are hierarchically ordered on the basis of learner control and level of learner interaction. For example, *browsing* offers the least learner control and is least interactive, because generally learners browsing through a multimedia environment lack specific intention or a defined goal, when *generating* allows students to go beyond controlling and sequencing the instruction to contribute to the instructional database.

While some learner control can motivate students, too much can be confusing (as cited in Litchfield, 1993). Learner characteristics such as prior knowledge and attitude can have a profound effect on knowledge acquisition in learner-controlled environments (as cited in Lawless et al., 1997). Snow (1980) indicates, "Learner control cannot be expected to overcome the persistent fact that individual characteristics not under the control of the individual will determine to a significant extent what and how much that individual will learn in a given instruction setting" (pp. 152-153). Hazen (1985) suggests that the optimal degree of learner control should be determined by learner characteristics, the nature of the content, and the complexity of the learning task.

Some students may not be able to make effective use of learner control. In order to help learners make decisions over their own learning and build effective learning strategies, literature (Jo, 1993) concludes the following recommendations for learner control to be integrated into the design of instruction:

- Control options should be clearly labeled to help learners use control options effectively.
- Immediate feedback, continuous advice on learners' on-going progress, and summaries of their uses of control options should be presented to help learners make "informed decisions" about their own learning.
- Basic requirement over important instructional components should be provided to learners in order to assure that they do not bypass the components.
- Prior to instruction, pre-training should be provided to learners to help them become familiar with the novel learning system with control options, perform conscious cognitive information processing, and understand objectives, procedures and values involved in building their own learning strategies (as cited in Jo, 1993).

CONCLUSION

Horng, Hong, Chanlin, Chang, and Chu (2005) reveal that multimedia use is one of the most important strategies of creative instructions. Multimedia not only has enormous potential for enhancing the learning, but also provides learners with varying levels of interactivity. To benefit from multimedia technology, teachers have to actively engage students in the presentation of information, and not just let the students become passive observers. Because teachers are the key to providing quality computer and interactive instructional media experiences for young children, the instructional materials they develop have to make the best use of multimedia technology within the framework of educational theories and learning principles. However, like Lawless et al. (1997) emphasize, "We have to be more cautious not to make the instructional system fit the technology but make the technology fit the instructional systems and formats that have been demonstrated to be effective. Technology is not effective learning in and of itself, but only provides a forum for effective learning" (p. 127-128).

Although today's multimedia technology provides exciting possibilities for creating quality learning experience and outcome, there will only be a minimal change impact if multimedia cannot be integrated with curriculum. Good teaching and real learning have thus become more important. Biggs (1999) states, "there is no single all-purpose best method of teaching. Teaching is individual" (p. 2). Multimedia can permit greater individualization if learning principles and strategies can be applied effectively. As the use of multimedia grows in education, multimedia pedagogical strategies will have a profound impact on how educators approach and engage students in the process of teaching and learning. This chapter discusses these principles and strategies for planning multimedia instruction, and intends to present them as guidelines for teachers using multimedia instructions in the classroom. The demand for making good use of multimedia to the best advantage for more effective and meaningful learning is continuously challenging educators teaching in such a rapidly evolving technological environment. Despite the dynamic nature of environment, educators' efforts of planning multimedia technology integration should be still ongoing and encouraging because the educator's role does not just stop at the planning multimedia instructions.

REFERENCES

Abram, S. (2005). The role of e-learning in the K-12 space. *MultiMedia & Internet@Schools, 12*(2), 19-21.

Agarwal, R., & Day, A. E. (1998). The impact of the internet on economic education. *Journal of Economic Education, 29*, 99-110.

Alavi, M., Yoo, Y., & Vogel, D. (1997). Using information technology to add value to management education. *Academy of Management Journal, 40*(6), 1310-1333.

Biggs, J. (1999). *Teaching for quality learning at university.* Buckingham: Open University Press.

Bitter, G., & Pierson, M. (2002). *Using technology in the classroom.* (5th ed.). Boston: Allyn & Bacon.

Chambers, B., Cheung, A., Gifford, R., Madden, N., Slavin, R. (2006). Achievement effects of embedded multimedia in a success for all reading program. *Journal of Educational Psychology, 98*(1), 232-237.

Chipman, S. F. (2003). Gazing yet again into the Silicon chip: The future of computers in education. In H. F. O'Neil, Jr., & R. S. Perez (Eds.), *Technology applications in education: A learning view* (pp. 31-54). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Christie, B., & Collyer, J. (2005). Audiences' judgments of speakers who use multimedia as a presentation aid: A contribution to training and assessment. *British Journal of Educational Technology, 36*(3), 477-499.

Conlon, T., & Simpson, M. (2003). Silicon Valley versus Silicon Glen: The impact of computers upon teaching and learning: A comparative study. *British Journal of Educational Technology, 34*(2), 137-150.

Doherty, P. B. (1998). *Learner control in asynchronous learning environments.* ALN Magazine, 2(2). Retrieved March 5, 2006, from http://www.sloan-c.org/publications/magazine/v2n2/doherty.asp

Ertmer, P. A. (1999). Addressing first-and second-order barriers to change: Strategies for technology integration. *Educational Technology Research & Development, 47*(4), 47-61.

Esquivel, G. B. (1995). Teacher behaviors that foster creativity. *Educational Psychology Review, 7,* 185-202.

Ewing, J. M., Dowling, J. D., & Coutts, N. (1998). Learning using the World Wide Web: A collaborative learning event. *Journal of Educational Hypermedia, 8*(1), 3-22.

Fahy, P. J. (2005). Planning for multimedia learning. In S. Mishra & R. C. Sharma (Ed.), *Interactive multimedia in education and training* (pp. 1-24). Hershey, PA: Idea Group Publishing.

Gayeski, D. M. (1993). *Multimedia for learning.* Englewood Cliffs, NJ: Educational Technology Publications.

Goldman, J. D. G., & Torrisi-Steele, G. (2005). Pedagogical design considerations in sex education on interactive multimedia using CD-Rom: An example of sexual intercourse. *Sex Education, 5*(2), 189-214.

Goldman, J. D. G., & Torrisi-Steele, G. (2002). Constructivist pedagogies of interactivity on a CD-Rom to enhance academic learning at a tertiary institution. *International Journal of Educational Technology, 3*(1), 1-27.

Gonzalez, R., Cranitch, G., & Jo, J. (2000). *Academic directions of multimedia education.* Retrieved February 28, 2006, from http://infotrac.galegroup.com/itw/infomark/423/684/81433982w5/purl=rc1_GBFM_0_A58615549&dyn=3!xrn_1_0_A58615549?sw_aep=mag_u_msu

Guan, S., & Mikolaj, P. (2002). *Collaborative problem solving in the online environment: A case study of a Web-based undergraduate business course.* (ERIC Document Reproduction Service No. ED477020)

Hannafin, M. J. (1989). Interaction strategies and emerging instructional technologies: Psychological perspectives. *Canadian Journal of Educational Communication, 18*(3), 167-179.

Hazen, M. (1985). Instructional software design principles. *Educational Technology, 25*(11), 18-23.

Holland, G. H. (2005). Streaming video: More than video through a computer. *MultiMedia & Internet@Schools, 41*(6), 6.

Horng, J., Hong, J., Chanlin, L., Chang, S., & Chu, H. (2005). Creative teachers and creative teaching strategies. *International Journal of Consumer Studies, 29*(4), 325-358.

Hoyles, C. (1994). Group work with computers: An overview of findings. *Journal of Computer Assisted Learning, 10*(4), 202-15.

Ivers, K. S., & Barron, A. E. (2002). *Multimedia projects in education. Designing, producing, and assessing.* Englewood, CO.: Libraries Unlimited, Inc. & Its Division, Teacher Ideas Press.

Jo, M. L. (1993). Hints and learner control for metacognitive strategies in problem solving. (ERIC Document Reproduction Service No. ED362169)

Jonassen, D. H., Howland, J., Moore, J., & Marra, R. (2003). *Learning to solve problems with technology: A constructivist perspective.* Upper Saddle River, NJ: Prentice-Hall.

Kaufman, C. (2002). *Reshaping curricular culture.* (ERIC Document Reproduction Service No. ED477038)

Kinzie, M. B. (1990). Requirements and benefits of effective interaction instruction: Learner control, self-regulation, and continuing motivation. *Educational Technology Research and Development, 38*(1), 5-21.

Kovalchick, A., & Dawson, K. (2004). *Education and technology: An encyclopedia.* California, Colorado, & England: ABD-CLIO, Inc.

Lawless, K. A., & Brown, S. W. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science, 25,* 117-131.

Leidner, D. E., & Jarvenpaa, S. L. (1995). The use of information technology to enhance management school education: A theoretical view. *MIS Quarterly, 19*(3), 265-291.

Lepper, M. (1985). Microcomputers in education: Motivational and social issues. *American Psychologist, 40*, 1-18.

Litchfield, B. C. (1993). *Design factors in multimedia environments: Research findings and implications for instructional design*. (ERIC Document Reproduction Service No. ED363268)

Marlow, E. (2003). *Problems in multi-media use in the reading curriculum*. (ERIC Document Reproduction Service No. ED479486)

Mayer, R. E. (2003). Theories of learning and their application to technology. In H. F. O'Neil, Jr. & R. S. Perez (Eds.), *Technology applications in education: A learning view* (pp. 127-158). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge: Cambridge University Press.

Means, B., & Haertel, G. D. (2004). *Using technology evaluation* to *enhance student learning.* New York & London: Teachers College Press.

Mevarech, Z. R., & Light, P. H. (1992). Cooperative learning with computers. *Learning and Instruction, 2*(3), 155-285.

Roblyer, M. D., & Schwier, R. A. (2003). *Integrating educational technology into teaching.* Toronto: Pearson Education Canada Inc.

Ross, S. M., & Morrison, G. R. (1988). Adapting instruction to learner performance and background variables. In D. H. Jonassen (Ed.), *Instructional designs for microcomputer courseware* (pp. 227-245). Hillsdale: Erlbaum.

Sanders, D. W., & Morrison-Shetlar, A. (2001). Student attitudes toward Web-enhanced instruction in an introductory biology course. *Journal of Research on Computing in Education, 33*(13), 25.

Savery, J. R., & Duffy, T. M. (1996). Problem based learning: An instructional model and its constructivist framework. In B. Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design*. Englewood Cliffs, NJ: Educational Technology Publications.

Schwier, R. A. (1993). Learning environments and interaction for emerging technologies: Implications for learner control and practice. *Canadian Journal of Educational Communication, 22*(3), 163-176.

Sims, R., & Hedberg, J. (1995). Dimensions of learner control: A reappraisal for interactive multimedia instruction. Retrieved March 10, 2006, from http://www.ascilite.org.au/conferences/melbourne95/smtu/papers/sims.pdf#search='dimensions%20of%20learner%20control'

Snow, R. E. (1980). Aptitude, learner control, and adaptive instruction. *Educational Psychologist, 15*, 151-158.

Speaker, K. (2004). Student perspectives: Expectations of multimedia technology in a college literature class. *Reading Improvement, 41*(4), 241.

Stone, L. (1999). Multimedia instruction methods. *Journal of Economic Education, 30*(13), 265.

Tearle, P., Dillon, P., & Davis, N. (1999). Use of information technology by English university teachers. Developments and trends at the time of the national inquiry into higher education. *Journal of Further and Higher Education, 23*(1), 5-15.

Tetreault, D. R. (2005). Administrative technology: New rules, new tools—A pilot study of excelsior software's electronic gradebook solution reveals the impact and time-saving qualities of administrative software. *T.H.E. Journal, 32*(9), 39.

Torrisi-Steele, G. (2005). Toward effective use of multimedia technologies in education. In S. Mishra & R. C. Sharma (Ed.), *Interactive multimedia in education and training* (pp. 25-46). Hershey, PA: Idea Group Publishing.

Vogel, D., & Klassen, J. (2001). Technology-supported learning: Status, issues, and trends. *Journal of Computer Assisted Learning, 17,* 104-114.

Wang, L., & Speaker, R. (2002). *Investigating education faculty's perspectives of their experiences in a technology project: Issues and problems related to technology integration.* (ERIC Document Reproduction Service No. ED477104)

Wild, M. (1996). *Perspectives on the place of educational theory in multimedia.* (ERIC Document Reproduction Service No. ED396746)

Yu. C., & Smith, M. (in press). PowerPoint: Is it an answer to interactive classrooms? *The International Journal of Instructional Media.*

Zazelenchuk, T. W. (1997). Interactivity in multimedia: Reconsidering our perspective. *Canadian Journal of Educational Communication, 26*(2), 75-86.

Zhang, D. (2005). Interactive multimedia-based e-learning: A study of effectiveness. *The American Journal of Distance Education, 19*(3), 149-162.

Zhao, Y. (2003). *What should teachers know about technology: Perspectives and practices.* Greenwich, CT: Information Age Publishing.

KEY TERMS

Asynchronous: A method of two-way transmitting data in which the parties present in the different time and space. An example of asynchronous communication is e-mail.

Computer-Assisted Instruction (CAI): Primarily refer to the use of computer(s) to present instruction to students. CAI is designed to help students learn new materials through interacting with the computer and students can progress learning with their own speed.

E-Learning: E-Learning is the use of network technology (broadly, the "Internet") to design, deliver, select, administer, and extend learning. Components of Internet-enabled learning can include content delivery in multiple formats, management of the learning experience, and a networked community of learners, content developers and experts.

Multimedia: The use of innovated technology to integrate text, graphics, animation, video and audio to transmit information.

Multimedia Instruction: Computer-based guidance that involves the use of diverse types of media, such as presentations, Web-based guides and online tutorials, in order to convey an instructional message.

Simulation: An interactive multimedia application device intended to imitate a real life situation and permit the user to partake and experience in a risk-free environment.

Synchronous: A method of two-way transmitting data in which the parties present in the same time and space. An example of synchronous communication is a chat room.

Virtual Reality: An interactive computer-based technology that allows the user to execute/perform actions in multi-dimensional setting.

This work was previously published in Handbook of Research on Instructional Systems and Technology, edited by T. T. Kidd and H. Song, pp. 216-231, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.8
XML Music Notation Modelling for Multimedia: MPEG–SMR

Pierfrancesco Bellini
University of Florence, Italy

ABSTRACT

The evolution of information technology has changed the use of music representation and notation in software applications, transforming and extending them from a simple visual coding model for music scores into a tool for modelling music for computer programs and electronic devices in general (e.g., keyboards), to support the exploitation of the multimedia characteristics lying behind music notation and representation. The MPEG symbolic music representation (MPEG-SMR) is a new emerging standard for modelling music notation within the MPEG multimedia framework. MPEG-SMR provides an XML-based language to model most of the music notation in terms of the visual and audio aspects, as well as music score annotations. MPEG-SMR also provides a language to define the music score formatting rules, supporting personalisation for the score visual presentation, custom symbols, and control visual rendering of the common notation symbols.

INTRODUCTION

Music in multimedia applications and frameworks is often considered only for its audible dimensions, while neglecting the important issues on the representation of the symbolic aspects. This could be due to historical cultural effects, since many popular and earlier multimedia tools are built for entertainment applications, and not focused on education, preservation, or research purposes. Music notation is an abstraction of the music. Not all performers use notations, and music notations have many different styles and forms. Currently, multimedia tools frequently use simple symbolic representations of music to represent the production of sound/music – for example, notes produced by an instrument.

Notations for the representation of music symbols have been developed over the years and ages to visually represent the pieces of information needed by a performer to play the music piece and reproduce the music as the author/composer

intended. The production of music notation scores for professional publishing on paper is one of the most traditional applications of music notation on computers (Blostein & Haken, 1991; Rader, 1996; Selfridge-Field, 1997).

The evolution of multimedia applications is accelerating relevant changes in the usages of music representation and notation in computer-based applications. Nowadays, it is no longer unusual to see music notation and modelling integrated into professional and educational music/audio applications (Bellini & Nesi, 2004; Byrd, 1984). In the past, several XML-based languages for music modelling have been proposed, including MNML (Musical Notation Markup Language), MusicML, MML (Music Markup Language), MusicXML (Good, 2001), WEDELMUSIC (http://www.wedelmusic.org) (Bellini & Nesi, 2001; Bellini, Della Santa, & Nesi, 2001), CAPXML (Capella, 2005), and so forth. Past efforts for music notation standardization were SMDL (SMDL, 1995) and NIFF (NIFF, 2005). Most of them are mainly focused on modelling the music elements to preserve and interchange the notation format and information among different applications (for editing and rendering of music scores), rather than to provide features that could support the integration of music notation with multimedia, for example, synchronisation with audiovisual and 3-D rendering, references and hyperlinks, multilingual lyrics, automatic formatting and rendering, and so forth. These features are clearly required and can be seen in tools from industrial projects, and R&D areas:

- Multimedia music for music tuition, such as VOYETRA, SMARTSCORE, PLAYPRO, MUSICALIS.
- Multimedia music for edutainment and infotainment, such as WEDELMUSIC (integrating music notation and multimedia to build and distribute multimedia-music cultural content with digital rights management), or to produce multimedia content for theatres: OPENDRAMA (http://www.iua.upf.es/mtg/opendrama/);
- Cooperative music editing, such as in MOODS (http://www.dsi.unifi.it/~moods), (Bellini, Fioravanti, & Nesi, 1999; Bellini, Nesi, & Spinu, 2002), and more recently using MAX/MSP with I-MAESTRO project (http://www.i-maestro.org).

Most of the applications mentioned are based on a multimedia music content format that is specific for each product. This is why any information exchange among the products can be so difficult, and it is strongly restricted to subsets of the notational part, for example, in MIDI. The lack of standardized symbolic music representation integrated with multimedia content results in each developer/company implementing their own solution, which may vary in efficiency, scope, features, quality, and complexity.

In this context, the MUSICNETWORK (http://www.interactivemusicnetwork.org) project began in 2002 to support a group of experts to identify a standard format for music representation for multimedia applications. The MUSICNETWORK started to work with ISO MPEG on the SMR (symbolic music representation), as described in another chapter of this book. The integration of SMR in MPEG multimedia framework, with technologies ranging from video, audio, interactivity, and digital rights management, has enabled the development of many new applications like those mentioned earlier and in Bellini, Nesi and Zoia (2005).

An overview of the MPEG-SMR standard is presented in this chapter.

MPEG SYMBOLIC MUSIC REPRESENTATION

The MPEG symbolic music representation (SMR), as specified in ISO/IEC 14496-23, is composed of three different languages:

- The Symbolic Music Extensible Format (SM-XF) to encode main score, single parts, and lyrics.
- The Symbolic Music Formatting Language (SM-FL) to customize music formatting style.
- The Symbolic Music Synchronization Information (SM-SI) that is used to provide synchronization information with the multimedia scene.

Figure 1 shows the relationships among the SMR data in the event of a music score with three parts, and how the SMR data is used by an MPEG-4 player to produce a synchronized multimedia presentation. Both SM-XF and SM-FL are XML languages defined using XML schemas, while SM-SI is binary encoded.

Interactivity features that may be implemented by an SMR-enabled MPEG-4 player are specified in an amendment to the BIFS (BInary Format for Scene representation) specification (ISO/IEC 14496-11:2005 AMD5), which is used to describe the multimedia scene.

SM-XF: SYMBOLIC MUSIC EXTENSIBLE FORMAT

Symbolic music extensible format is an XML application for encoding main scores, single parts, as well as multilingual lyrics. In Figure 2, a simplified UML diagram of the SMR model is presented. At an abstract level, the main score consists of single parts, each single part consists of a sequence of measures, and each measure consists of parallel layers containing sequences of notes, rests, and other timed and untimed symbols that are organized horizontally on a staff. Each symbol (e.g., notes, rests, chords, etc.) can be associated with some qualifier symbols to represent some additional pieces of information on the symbol itself, such as expression symbols. A chord is modelled as a sequence of chord notes (one for each note head), and beamed notes are modelled as containers of musical figures. Additionally, there are other symbols, called horizontal symbols, spanning over multiple timed symbols (e.g., slurs, crescendo, diminuendo), that start at a specific musical figure and end at another musical figure. The horizontal symbols are contained

Figure 1. Example of structure and relationship among MPEG-SMR data and how a SMR enabled MPEG-4 player should use it

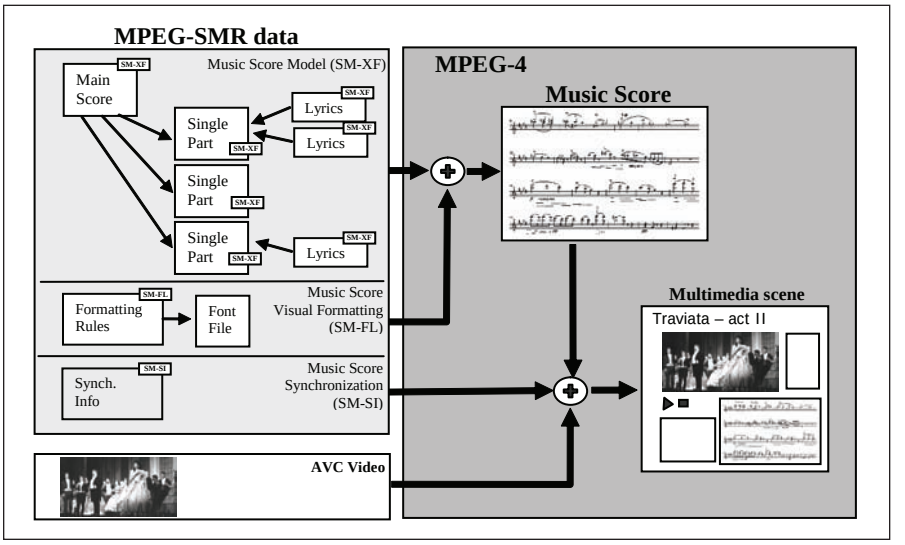
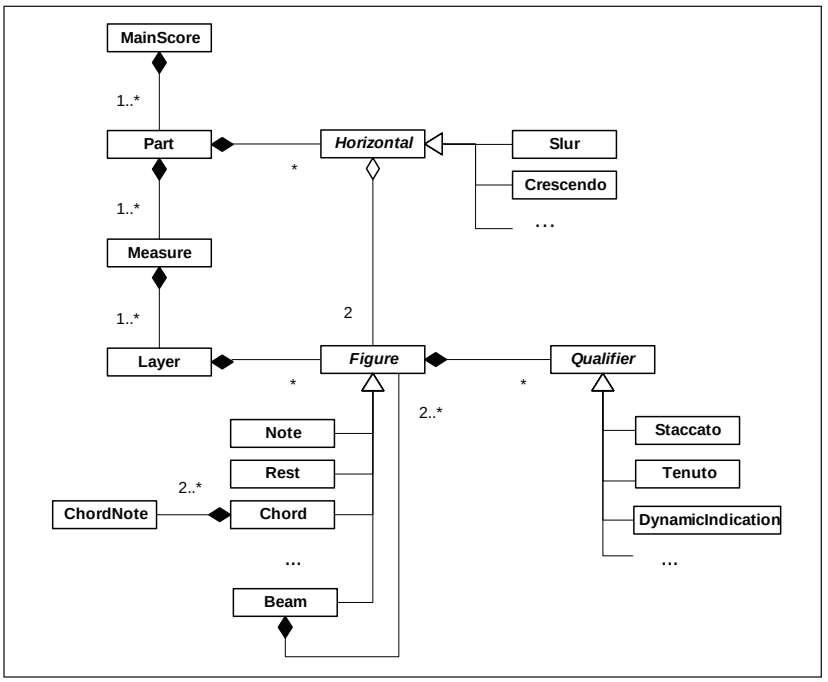


Figure 2. SMR abstract UML model



inside the single part and referred to the start/end figures/locations.

The MainScore representation is built from the Parts representation. Other formats use different representations for MainScore and Parts; thus, having a strong duplication (each part is provided twice, in the main score and in the single instrumental part). If the main score is built from parts, it reduces the size of the data (to be sent to the client) and supports a synchronized visualization of parts and main score (a modification made on a part in a main score view is reproduced mirror-like in the part shown in the single part view). In some cases, the main score and the parts are encoded differently, and some musical elements are present, specifically for the view (e.g., small guideline indications are used in the single parts) or different instruments are merged on a single staff in the main score representation. In such cases, the representation can be duplicated, and the main score and parts can be represented separately.

In SMR, a measure is represented as a container of musical figures. This is different from other formats such as CapXML (Capella, 2005), which does not group musical figures in measures and have the barline as another symbol among musical figures. However, the structural subdivision in measures enables subdivision of the content in chunks to be delivered to clients. In MPEG-4, all the media content (and subsequently also SMR) is delivered to clients in subdivided Access Units. The Decoder receives the Access Units one after another and it has to decode them, passing the decoded data to the renderer. In SMR, each Access Unit delivered to the terminals can contain one or more measures for the different parts. The whole score can be provided in one Access Unit or subdivided in many Access Units, thus enabling the device to start rendering the score without being restricted to wait for the whole score (consider the case of an opera containing hundreds or thousands of measures).

One of the main requirements in the language design consisted in each element (Part, Measure,

Layer, Note, Rest, Chord, ChordNote, …) being identified in a unique way to allow musical elements to be referred directly in the score.

This feature is used in:

- Horizontal symbols (e.g., slurs, dynamics) to indicate the start/end element of the symbol.
- Lyrics to indicate the note a syllable is connected to.
- Annotations to indicate to which musical symbols the annotation is related/associated to.

In all these occurrences, the information is stored outside the score itself. This feature allows a primary musical structure to associate additional information, such as annotations. Moreover, the musical element identification has to work, even if it does not have in memory the whole score (this is particularly useful for low-memory footprint devices), and the identifier has to be valid even after score manipulation (therefore, element position cannot be used as identifier). For these reasons, each element has been identified with a numeric ID that is unique in the parent element (Part in Main Score, Measure in Part, Layer in Measure, Figure in Layer, ChordNote in Chord). Hence, an element can be identified by using a sequence of IDs that specify the path to be followed in order to locate the element. The path is valid even if new measures/figures are added before the identified element. If each element is easily identified, this allows for things such as separate files for Lyrics or Annotations that can be applied without any modifications to the score.

Beams are modelled as containers of Figures inside Layers, thus not allowing beaming across measures. Hence, a kind of horizontal symbol used to beam notes from a start note to an end note has been introduced. In this case, the end note can be in another measure, to support beams across measures. This feature has been added only recently and both approaches can be used.

For instruments using more than one staff (like piano and organ), the staves are considered as belonging to the same score (other XML formats, like CapXML, encode each staff separately). The score has an attribute stating how many staves are used by the instrument, and each figure in a layer has an attribute indicating which staff the note/rest belongs to. Therefore, layers can go from one staff to another and can easily represent beaming across staff.

What follows is a set of selected examples on how the basic musical symbols are represented in SMR, starting from the bottom level (a note) and going up to a single part and a main score. The complete description of the XML language can be found in the ISO/IEC MPEG Specification (ISO/IEC 14496-23 FDIS).

In Figure 3, a brief explanation of the notation used to represent XML elements is reported.

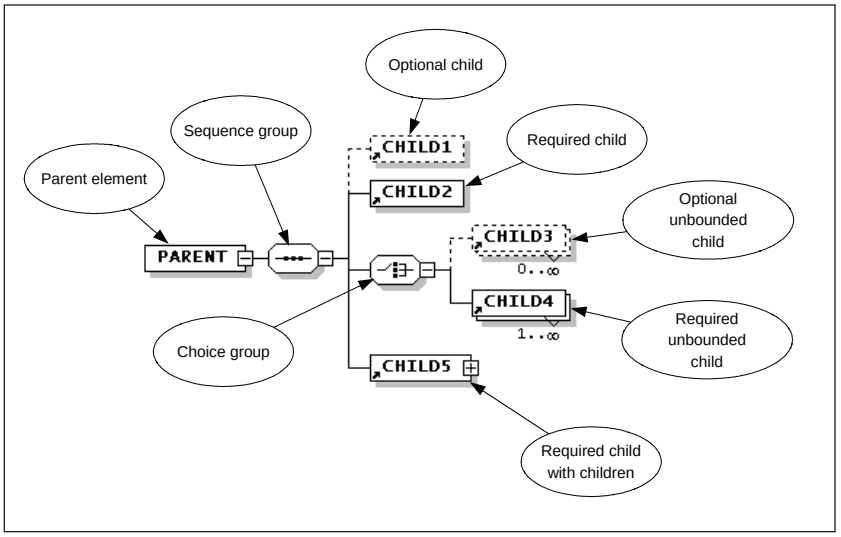
Note

The musical note is modelled in XML with the note element, as described in Figure 4 (attributes documentation is missing). A note element may contain:

- An optional pitch element with the note pitch indication.
- Accidental symbols (like sharp, flat, etc.).
- Augmentation information that modifies the note duration.
- A sequence of qualifier symbols like fermata, dynamic text, markers, and so forth.
- An optional user spacing with the distance with respect to the next figure.

If the pitch is missing, it can be deduced from the HEIGHT attribute containing the position of the note on the staff, from the accidental symbols, if any, and from the clef and key signature currently active. If both the height attribute and the pitch element exist, the height is used for visual rendering and the pitch information is used for

Figure 3. Notation used to represent XML element structure



audio rendering. In the event that the height attribute is not specified, the latter is deduced from the pitch and from the clef and key signature.

Some of the attributes of the note element are:

- The HEIGHT containing the position of the note on the staff (0 for the lowest staff line, 1 for the first space, 2 for the second staff line, etc.).
- The DURATION with the note duration (e.g., D1_8 for an eighth note).
- The ID, which identifies the note within the layer.
- The STEM to indicate the direction of the stem (up or down).
- The STAFF to indicate the staff where the note has to be positioned.

Figure 4 (on the right side) reports an example on how a note is represented in XML.

Rest

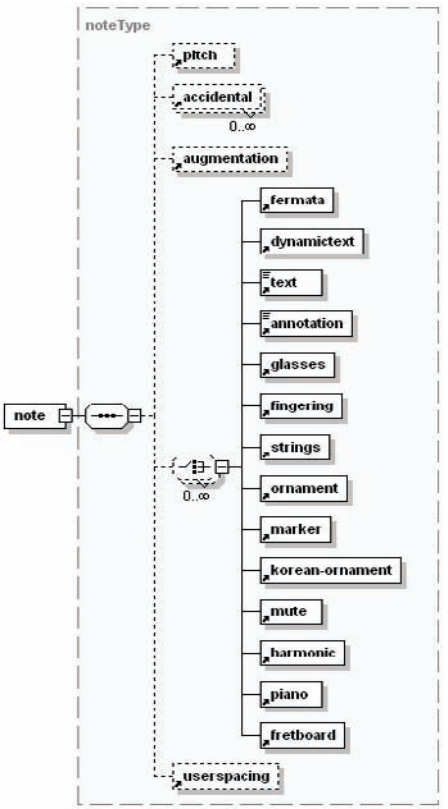
A rest symbol is represented in XML with the rest element, which may contain the augmentation information so as to change the rest duration. It may contain zero or more qualifier symbols, such as fermata, dynamic text, textual indication, annotation, pay attention symbol (glasses), piano symbol, and fretboard symbol. Moreover, it may contain the userspacing element expressing the distance from the next figure.

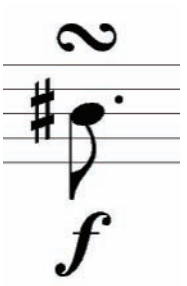
The rest element contains attributes such as the rest DURATION, the HEIGHT with the position of the rest on the staff, and the rest ID to identify the rest in the layer.

Chord

A chord is represented in XML with the chord element containing a sequence of chordnote elements, with the information related to each note head of the chord (accidentals, fingering, pitch, position on the staff, staff, …). The chord element contains the same elements of the note (fermata, dynamics, text, etc.), since they refer to the chord as a whole. The arpeggio element is specific for chords to indicate how to play the chord. Figure 6 depicts the XML element structure with an example of a chord.

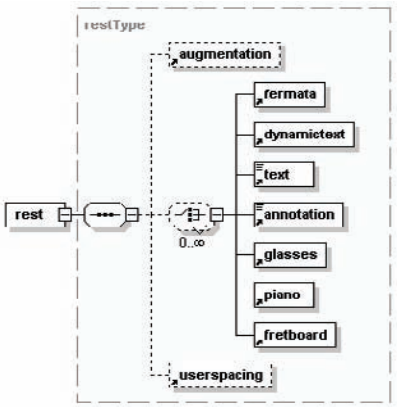
Figure 4. XML model of a note (on the left) and an example (on the right)





```
<note ID="1" DURATION="D1_8" STEM="DOWN" HEIGHT="4">
   <pitch BASE="B" SHIFT="#" OCTAVE="4"/>
   <accidental TYPE="SHARP" ROUNDBRACKET="NO"/>
   <augmentation DOTS="1"/>
   <dynamictext DYNAMIC="F" UPDOWN="DOWN"/>
   <ornament TYPE="TURNBACK" UPDOWN="UP"/>
</note>
```

Figure 5. XML model of a rest (on the left) and an example (on the right)



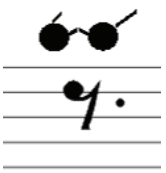
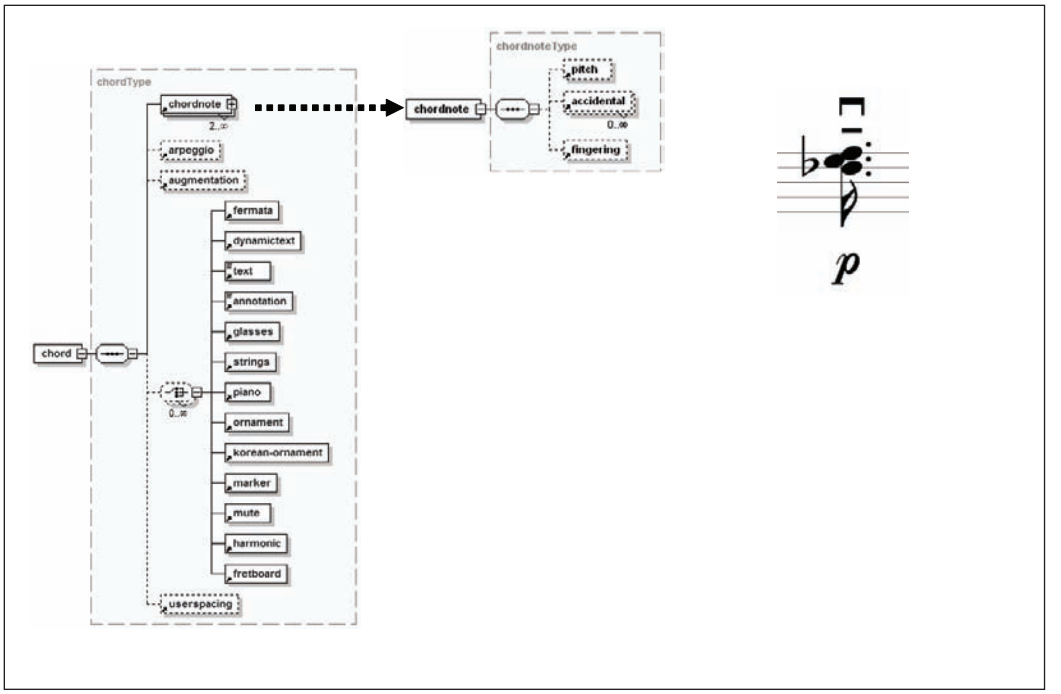


Figure 6. XML model of a chord (on the left) and an example (on the right)



Beam

The beam is modelled as a container of notes, chords, rests, anchorages, and change clefs elements. The anchorage element represents a point in the score where other symbols (mainly horizontal symbols) can be logically attached. With this kind of model, a beam crossing the bar line cannot be represented; for this reason a specific horizontal symbol can be used (see later on). Therefore, a beam can be represented both as container and horizontal.

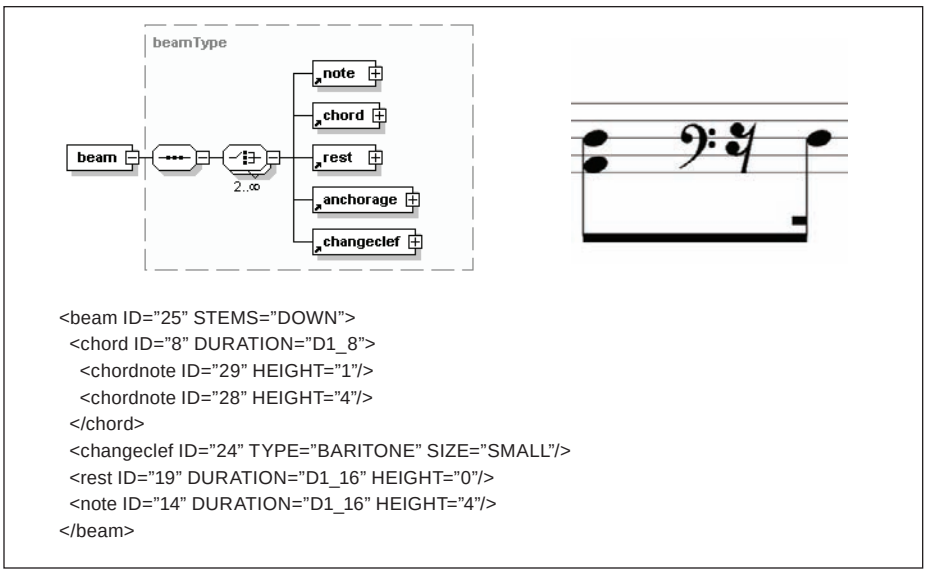
Measure

The measure represents the classical subdivision of a score; it is modelled with the measure element, which contains a sequence of layers containing the notes, rests, chords, beams, and so forth.

With some further details, the measure may contain:

- Justification information on how to position the notes/rests on the available space (e.g., linear or logarithmic with a tuning parameter).
- An optional label for the measure (e.g., rehearsal marks or "segno/coda" signs).
- An optional jump indication to indicate the successive measure to be executed (e.g., da capo al fine).
- A header with the clef and key signature for each staff.
- The time signature of the measure (e.g., 3/4).
- A beat scan indication.
- An optional metronome indication that applies starting from the measure until another metronome change.
- The different layers with the musical figures.
- The bar line to be used (e.g., single, double, refrain start/end, invisible).

Figure 7. XML model of a beam with an example




```
<beam ID=”25” STEMS=”DOWN”>
 <chord ID=”8” DURATION=”D1_8”>
  <chordnote ID=”29” HEIGHT=”1”/>
  <chordnote ID=”28” HEIGHT=”4”/>
 </chord>
 <changeclef ID=”24” TYPE=”BARITONE” SIZE=”SMALL”/>
 <rest ID=”19” DURATION=”D1_16” HEIGHT=”0”/>
 <note ID=”14” DURATION=”D1_16” HEIGHT=”4”/>
</beam>
```

Figure 8. XML model of the measure

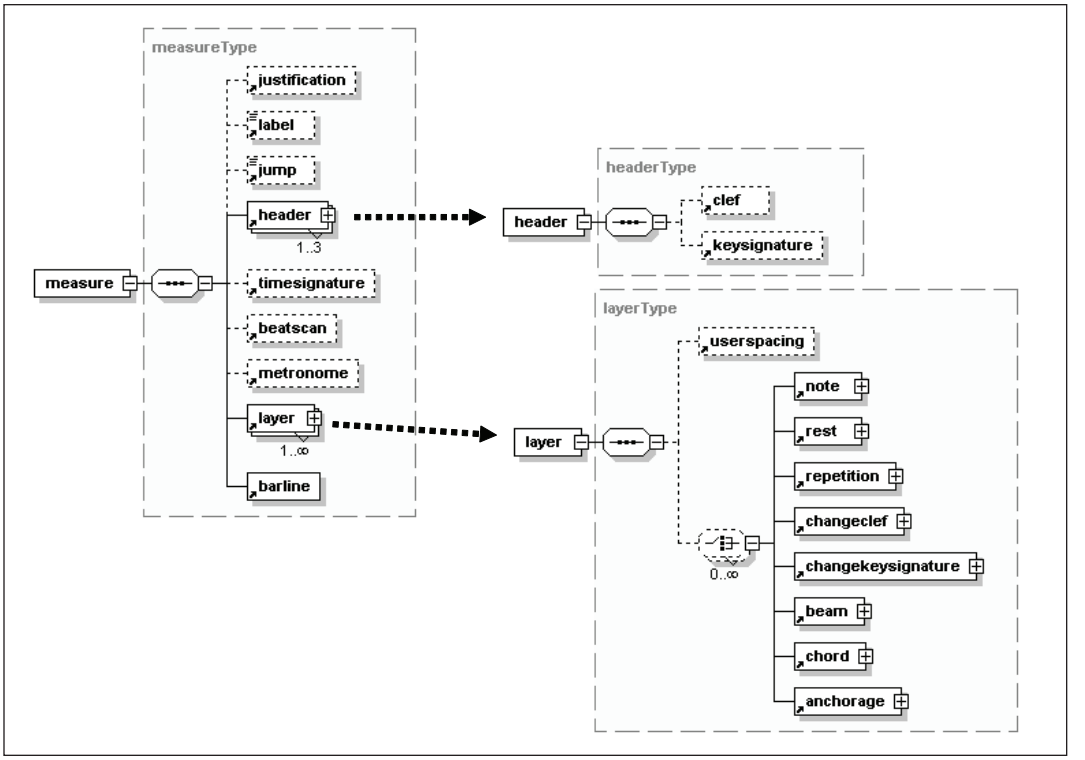
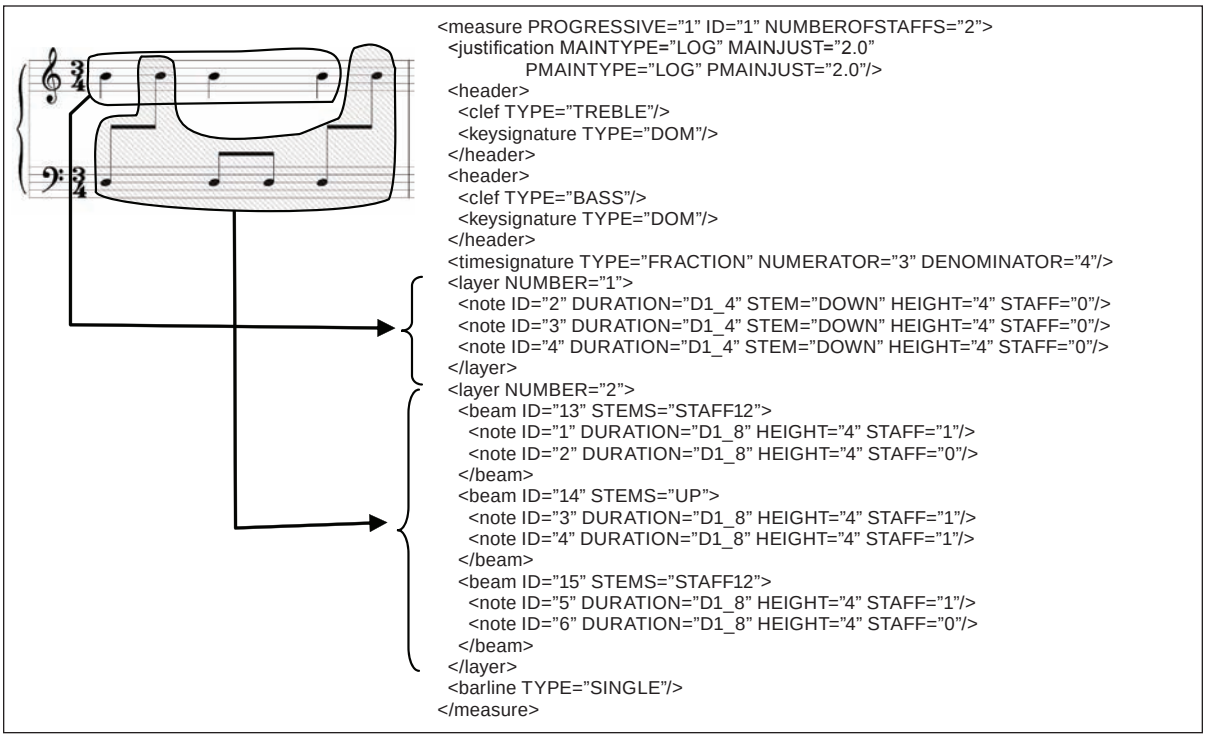


Figure 9. An example of measure



```
<measure PROGRESSIVE="1" ID="1" NUMBEROFSTAFFS="2">
 <justification MAINTYPE="LOG" MAINJUST="2.0"
          PMAINTYPE="LOG" PMAINJUST="2.0"/>
 <header>
  <clef TYPE="TREBLE"/>
  <keysignature TYPE="DOM"/>
 </header>
 <header>
  <clef TYPE="BASS"/>
  <keysignature TYPE="DOM"/>
 </header>
 <timesignature TYPE="FRACTION" NUMERATOR="3" DENOMINATOR="4"/>
 <layer NUMBER="1">
  <note ID="2" DURATION="D1_4" STEM="DOWN" HEIGHT="4" STAFF="0"/>
  <note ID="3" DURATION="D1_4" STEM="DOWN" HEIGHT="4" STAFF="0"/>
  <note ID="4" DURATION="D1_4" STEM="DOWN" HEIGHT="4" STAFF="0"/>
 </layer>
 <layer NUMBER="2">
  <beam ID="13" STEMS="STAFF12">
   <note ID="1" DURATION="D1_8" HEIGHT="4" STAFF="1"/>
   <note ID="2" DURATION="D1_8" HEIGHT="4" STAFF="0"/>
  </beam>
  <beam ID="14" STEMS="UP">
   <note ID="3" DURATION="D1_8" HEIGHT="4" STAFF="1"/>
   <note ID="4" DURATION="D1_8" HEIGHT="4" STAFF="1"/>
  </beam>
  <beam ID="15" STEMS="STAFF12">
   <note ID="5" DURATION="D1_8" HEIGHT="4" STAFF="1"/>
   <note ID="6" DURATION="D1_8" HEIGHT="4" STAFF="0"/>
  </beam>
 </layer>
 <barline TYPE="SINGLE"/>
</measure>
```

Some of the attributes of the measure element are:

- An ID to identify the measure in the score.
- A progressive number;
- The number of staves to be used (1, 2 for piano, or 3 for organ).

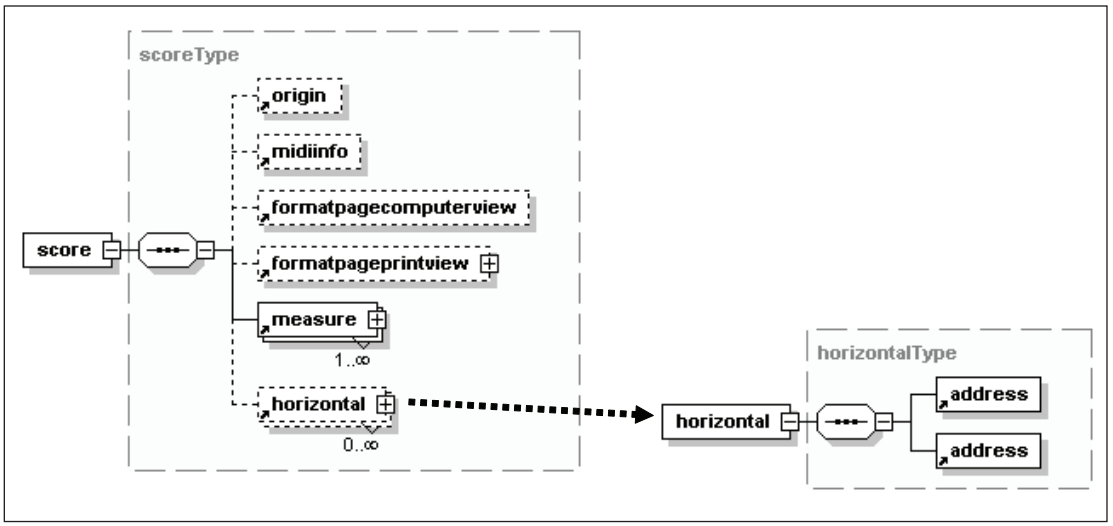
Figure 9 provides an example of a measure spanning on two staves. Please note that two layers are used, one for the notes on the upper staff and one for the notes on the lower staff, and in the second layer there are two beams with notes belonging to different staves.

Score and Horizontal Symbols

The score is modelled as a container of both a sequence of measures and horizontal symbols. Horizontal symbols are the ones spanning over the score from a start event/position up to an end event/position (e.g., slurs, crescendo/diminuendo). The start/end event/position may correspond to a musical figure, such as a note, rest, or chord, or to an anchorage point that represents an event between two musical figures. Since horizontal symbols are not stored within the measures but outside, what is needed is a way to logically identify the start/end element of the symbol. In order to refer to the start/end figure of the score, a set of identifiers is used in the address element:

- The ID of the measure where the start/end figure is.
- The number of the layer where the figure is.
- The ID of the beam/chord/note/rest within the layer containing the addressed figure.
- The optional ID of the note within the chord or the ID of the note/rest/chord within the beam.

Figure 10. The XML model of the score element



- The optional ID of the note within the chord in a beam.

Horizontal symbols are used for:

- Slurs
- Ties
- Tuplets (e.g., terzine)
- Octave change (8va, 8ba, 15ma, 15ba)
- Beaming across measures
- Crescendo/diminuendo
- Trill with a wavy line
- Bend
- A wavy line
- An arrow
- Refrain change
- Piano pedal indication

The score is represented in XML as illustrated in Figure 10. The score element can contain:

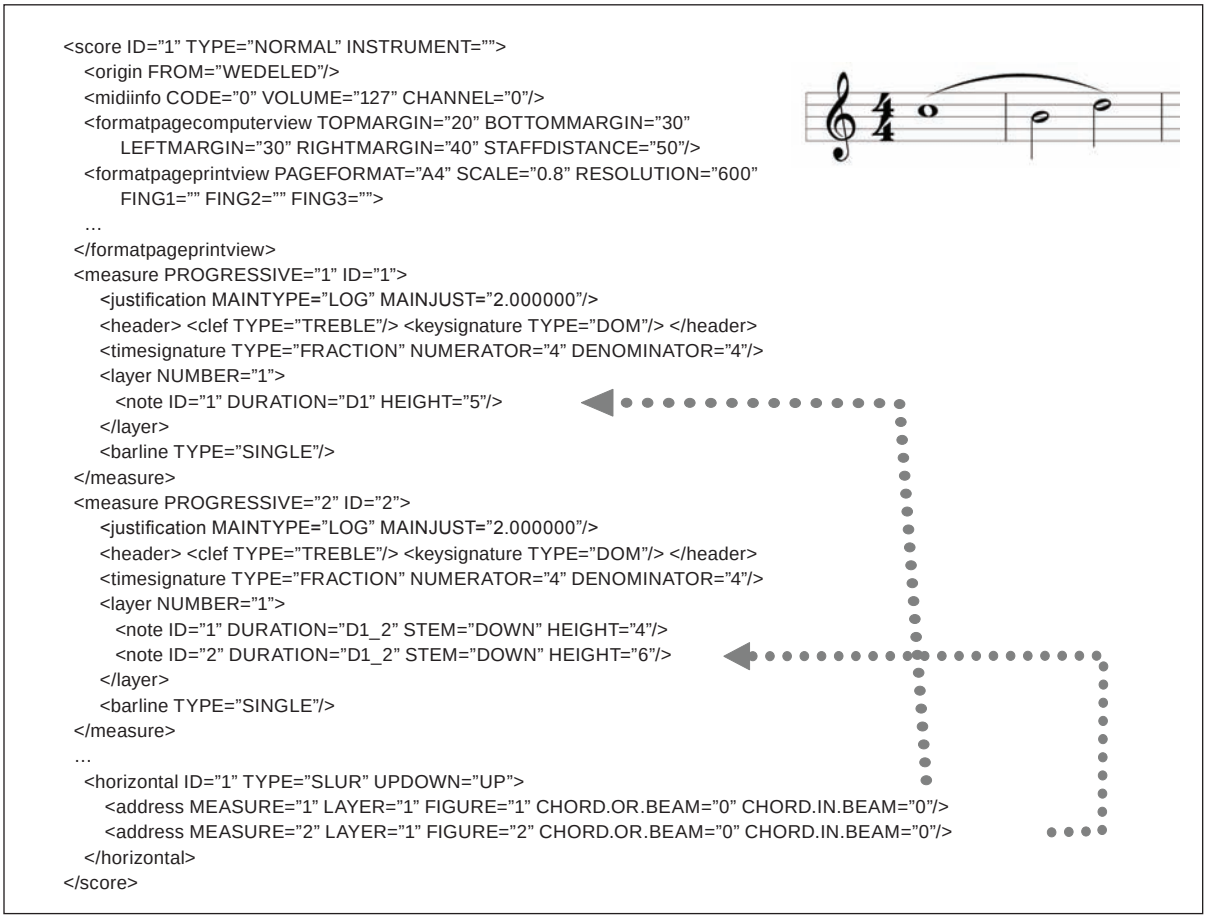
- Some optional origin information (e.g., the software tool used to produce the score).
- Some optional MIDI information on the instrument, volume, and channel to be used for MIDI execution.
- Some optional information on how to format the page for computer view and for print view (e.g., margins, distance between staves, number of systems per page).
- The sequence of measures building the score.
- The sequence of horizontal symbols, each one with two addresses identifying the start/end figure/event.

Attributes of the score element include:

- The score ID identifying the score
- The score type (e.g., normal, percussion, tablature)
- The instrument name
- The number of staves to be used (e.g., 1, 2 or 3)
- …

Figure 11 presents an example of a score with a slur connecting two notes.

Figure 11. An example of an MPEG-SMR score



Single-Part Score

The single part contains the musical information for one executor, and it is modelled with the SMXF_Part element. It contains the score element with the musical information and some general identification and classification information (as XML elements), some preferences for some specific decoder, and additional pieces of information for printing the score as a single part. The printpages element has textbox and imagebox elements, to be used when printing each page.

Main Score

The main score contains the musical information of the whole score and subsequently, it contains the information about all the single parts of the instruments playing the music piece.

Like the single part, the main score may contain identification and classification information and custom preferences for specific decoders. Moreover, the main score contains:

- References to the single parts making the main score (using the score IDs).
- Some general MIDI information (e.g., how to map to MIDI dynamic symbols).
- Some formatting information for the computer view and for print view (e.g., margins, distance between staves, number of systems per page).
- A sequence of brackets used to group visually different parts (e.g., string instruments).

Figure 12. XML model for the single part

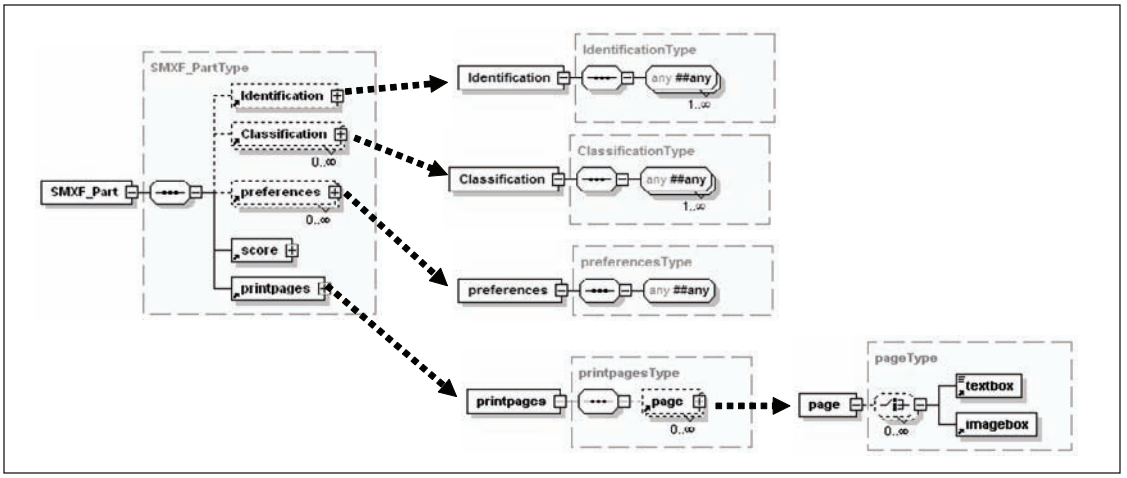


Figure 13. XML model of the main score

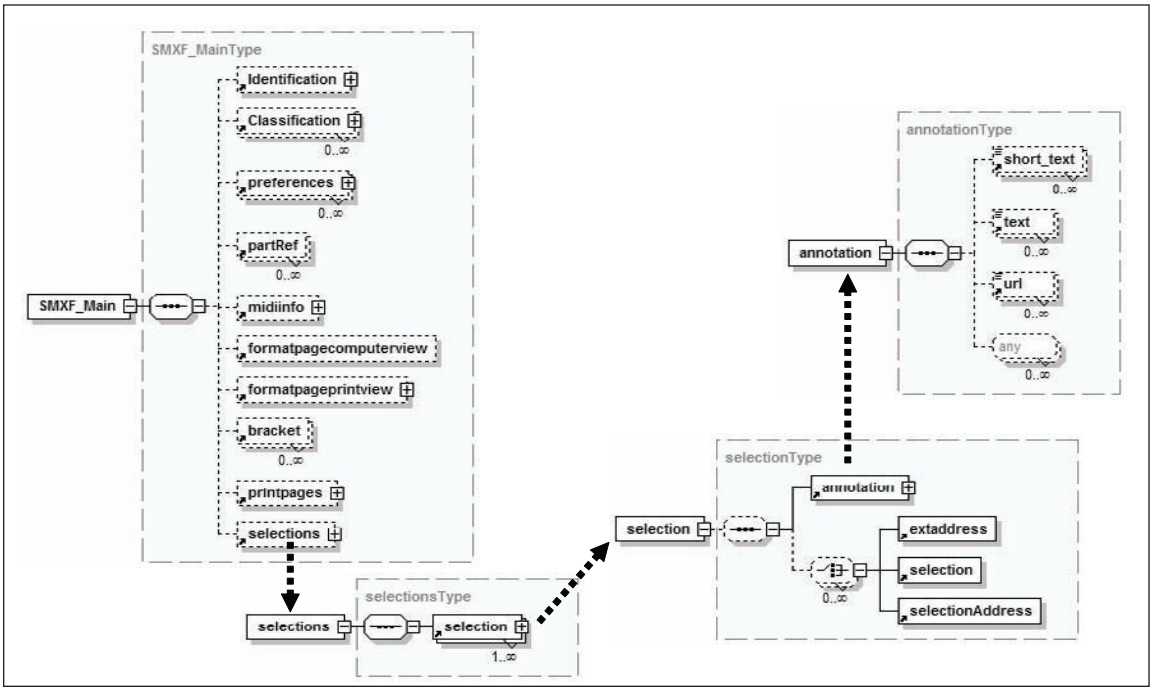
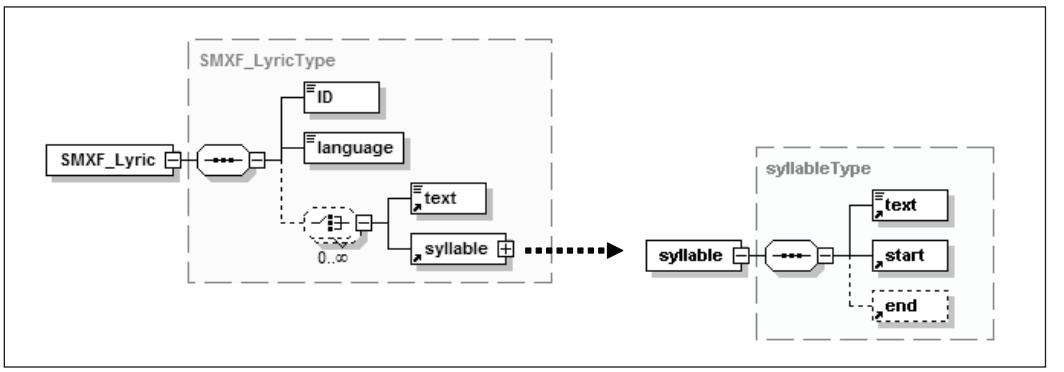


Figure 14. XML model for the lyrics



- Additional "graphical" elements (image and text boxes) for print view.
- An optional sequence of selections used to associate annotations (as text, URL, or any XML elements), with a subset of musical symbols of the different parts.

Selections are used to group together a subset of musical figures being included in the score, and to annotate them with textual descriptions (in different languages) or reference to a digital resource via URLs (e.g., an image contained in the MPEG-4 file), or with any customized XML content. A selection may be defined with extaddress elements that refer to the musical figures in the score (the extaddress contains the score ID that is not included in the address element of the horizontal symbols). Moreover, the selection can be built also by embedding other selections so as to form a hierarchical structure, or by referring to other selections using an ID (using the selectionAddress element).

Selections are useful in music education in order to mark specific passages or to give audio/visual performance indications.

Lyrics

The lyrics are modelled in a way similar to the horizontal symbols; the syllables of the lyrics are not spread over the notes in the score (like many other music notation formats) but are kept separate, and they refer to the notes they are associated with (using the same mechanisms of addresses, as it happens with the horizontal symbols). In this way, different lyrics (e.g., in different languages) can be plugged on a score without changing the score. Moreover, the original lyrics can be reconstructed as plain text, out of the lyrics model. This method to model lyrics is explained with further details in Bellini et al. (Bellini, Bruno, & Nesi, 2004).

The SMXF_Lyric element contains:

- An ID to identify the lyrics.
- The language used for the lyrics.
- A sequence of mixed text and syllable elements, where the text elements contain some accompanying text that should not appear in the score (e.g., some formatting text), and syllable elements containing information on the syllable.

The syllable element contains:

- The syllable text.
- A reference to the note the syllable is associated with (the start element).
- An optional end position in case of syllable prolongation like in melisma, where the same syllable is sung over more notes.
- An attribute with the row where the syllable has to be positioned, to be used in case of refrains.

In Figure 15, an example of lyrics is reported.

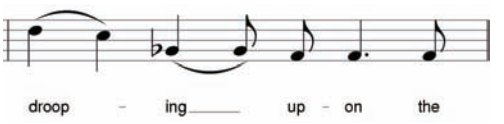
SM-FL: SMR FORMATTING LANGUAGE

The symbolic music formatting language (SM-FL) is defined to allow the description of the insertion point and the positioning of common music symbols (stem, ornaments, expression, etc.), as well as the definition of new symbols. It specifies a rule-based formatting language and engine that is used to describe sets of rules and conditions to be applied and interpreted whenever the position of symbols has to be estimated. The SM-FL rules define formatting actions; they assign specific values to specific parameters related to the visual rendering of the music symbols. For example, a rule is used to define the stem length of a note. The rule to be applied is identified on the basis of the conditions met in the music context of the symbol under evaluation. These conditions describe a music scenario. For example, a music scenario could consist of a note not belonging to a chord, that is, nonpolyphonic, and with a height set within a certain range. The verification of a music scenario defined by a conditional sentence leads to the application of a certain formatting rule to the symbol under evaluation.

SM-FL can be used to define rules to customize:

- Stem length and direction
- Beam slope and direction
- Automatic beaming rules
- Note head shape
- Stem start position
- Symbols' position and direction
- Symbols' positioning order (w.r.t. the note)
- Shape of any symbol

Figure 15. An example of MPEG-SMR SM-XF lyrics




```
<SMXF_Lyric SCOREID="1">
  <id>001</id>
  <language>en</language>
  …
  <syllable LINE="1" SEP="-">
        <text>droop</text>
        <start MEASURE="10" LAYER="1" FIGURE="1"/>
        <end MEASURE="10" LAYER="1" FIGURE="3"/>
  </syllable>
  <syllable LINE="1" SEP="_">
        <text>ing</text>
        <start MEASURE="10" LAYER="1" FIGURE="3"/>
        <end MEASURE="10" LAYER="1" FIGURE="4"/>
  </syllable>
  <syllable LINE="1" SEP="/">
        <text>up</text>
        <start MEASURE="10" LAYER="1" FIGURE="5"/>
        <end MEASURE="10" LAYER="1" FIGURE="7"/>
  </syllable>
  <syllable LINE="1" SEP=" ">
        <text>on</text>
        <start MEASURE="10" LAYER="1" FIGURE="7"/>
  </syllable>
  <syllable LINE="1" SEP=" ">
        <text>the</text>
        <start MEASURE="10" LAYER="1" FIGURE="8"/>
  </syllable>
  …
</SMXF_Lyric>
```

SM-FL can be used to define new symbols associated with a note and rules to cope with their position in the score. The SM-FL is an XML language derived from the MILLA language used in WEDELMUSIC editor (Bellini et al., 2005). In Figure 16, the structure of an SMFL rule file is reported, it contains:

- A sequence of font-mapping definition elements where the shape of some classical symbols (clefs, alterations, etc.) can be redefined (those represented using fonts).
- A sequence of group definitions allowing to define new groups containing custom symbols.
- In any order: rule definition elements to set a particular aspect of the score (note head shape, stem direction, stem length, beam direction and slope, etc.) and rule application elements defining the condition for a specific rule's application.

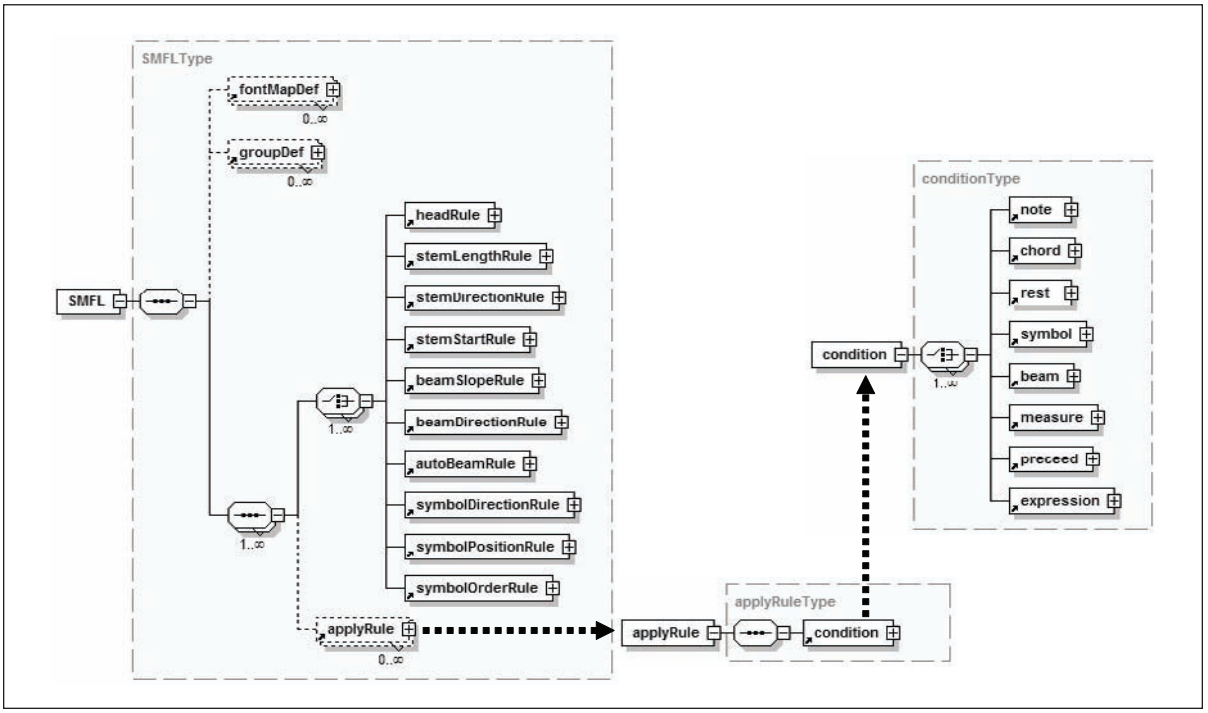
The only rule that is not conditional is the SymbolOrderRule, which is used to state the position of symbols with respect to the note head. In order to be applied, all the other rules need a condition to be satisfied.

For example, an insertion rule "StemUp," which sets the stem upward with respect to the note head, can be stated as:

```
<stemDirectionRule ruleId="StemUp">
        <stemUp/>
</stemDirectionRule>
```

A condition to activate this rule can be very simple. The condition could state that the rule "StemUp" is applied whenever the note is found below the middle line of the staff:

Figure 16. XML model for SM-FL



```
<applyRule rule=”StemUp”>
        <condition>
        <note>
                        <heightLT>0</heightLT>
        </note>
        <condition>
</stemDirectionRule>
```

A different condition may state that the rule “StemUp” is invoked if the note belongs to the upper voice for a measure having polyphony (In-MultivoiceUpper) as a feature. The upper voice is the one presenting the note with the highest pitch among those to be played at the same time:

```
<applyRule rule=”StemUp”>
    <condition>
        <note>
            <inMultivoiceUpper/>
        </note>
    <condition>
</applyRule>
```

Another case is when the note is in a single layer and is included in a chord:

```
<applyRule rule=”StemUp”>
  <condition>
    <note>
      <inSinglevoice/>
      <inChord/>
    </note>
    <expression>
      <lt>
        <minus><chord.upperd/><chord.lowerd/></minus>
        <value>0</value>
      </lt>
    </expression>
  <condition>
</applyRule>
```

Such conditions are met in the second measure of Figure 17. The notes belong to a chord (inChord), only one voice is present (inSinglevoice), and the difference between the highest and the lowest notes of the chord defines the “centre of gravity” of the chord, either above or below the middle line, that is (upperd-lowerd>0), where: upperd is the absolute value based on the distance between the highest note of the chord and the middle line of the staff, and lowerd is the absolute value based on the distance between the lowest note of the chord and the middle line of the staff.

Specific rules can be provided to set the stem length. The basic unit for stem length is the space defined as the distance between two staff lines. In this way, the standard length of the stem is 3.5 spaces, while it has to assume different values, depending on the note height. In Figure 18, the following rules and conditions have been used for some notes:

```
<stemLengthRule ruleId=”Stem3_5”>
    <length>3.5</length>
</stemLengthRule>
<stemLengthRule ruleId=”StemHeight”>
    <noteHeight/>
</stemLengthRule>

<applyRule rule=”Stem3_5”>
    <condition>
        <note>
            <heightGE>0</heightGE>
            <heightLE>7</heightLE>
            <stemDown/>
        </note>
    <condition>
</applyRule>
<applyRule rule=”StemHeight”>
    <condition>
        <note>
            <heightGE>8</heightGE>
            <stemDown/>
        </note>
    <condition>
</applyRule>
```

Figure 17. The stem of notes and chords in single layer and in polyphony

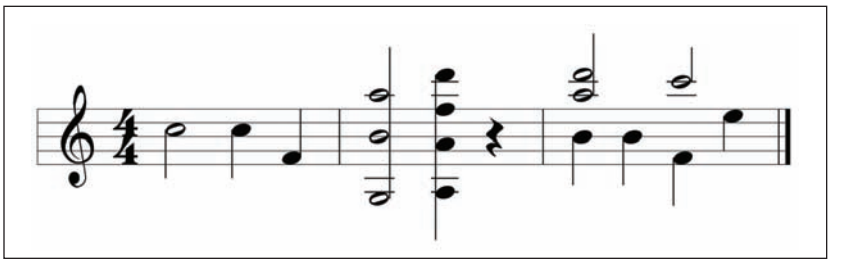


Figure 18. Example for stem length

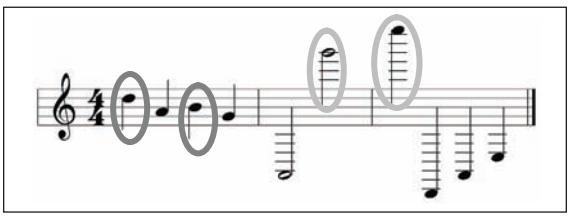
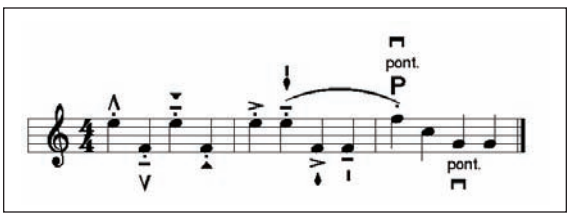


Figure 19. An example of order rule application



The first condition is verified for the first note of the first measure of Figure 18, while the second condition is true for the second note of the second measure and for the first of the third measure, which sets the use of a stem with a length equal to the note height divided by two.

Markers and other symbols may accompany the same note. In this case, the SymbolOrderRule defines the order of symbols' positioning.

To change the order rule is a procedure that can be used to change the order of symbols without modifying the symbolic description of music score. The order rule appears as a list of symbol identifiers shown in decreasing order of priority. For example:

```
<symbolOrderRule>
    <symbol>STAC</symbol>
    <symbol>TEN</symbol>
    <symbol>SLUR</symbol>
    <symbol>MARTF</symbol>
    <symbol>ACCE</symbol>
    <symbol>SFOR</symbol>
    <symbol>MART</symbol>
    <symbol>ARCO</symbol>
    <symbol>PUNTA</symbol>
    <symbol>TALLONE</symbol>
    <symbol>PONTICELLO</symbol>
    <symbol>TASTIERA</symbol>
    <symbol>ARCATASU</symbol>
    <symbol>ARCATAGIU</symbol>
    <symbol>CORDA</symbol>
    <symbol>PIZZICATO</symbol>
    ...
</symbolOrderRule>
```

Therefore, the staccato symbol (STAC, if any) is the symbol closest to the note head, the tenuto symbol (TEN) the second closest one, and so forth.

The SM-FL allows the definition of new symbols, which are considered as generic expression symbols related to a note. For these new symbols, rules can be defined. Since symbol rules are usually very similar, symbols are grouped and rules are defined for groups of symbols. However, specific rules can be defined as well.

A new symbol can be defined using the symbolDef element inside a groupDef element:

```
<groupDef name=”faces” font=”mysym.ttf”>
    <symbolDef name=”smile”>
        <code>33</code>
        <dimension>
                <toTop>20</toTop>
                <toBottom>20</toBottom>
                <toLeft>20</toLeft>
                <toRight>20</toRight>
                <dx>-20</dx>
                <dy>-20</dy>
            </dimension>
        </symbolDef>
        <symbolDef name=”sad”>
            <code>34</code>
            <dimension> … </dimension>
        </symbolDef>
        <symbolDef name=”star5”>
            <code>35</code>
            <dimension> … </dimension>
        </symbolDef>
</groupDef>
```

specifying:

- The name of the symbol (e.g., “smile”).
- The group it belongs to (e.g., “faces”).
- The name of the font where the symbol can be found (e.g., “mysym.ttf”).
- The code of the character representing symbol in the font file (e.g., code 36).
- The bounding box of the symbol in the dimension element (see Figure 20 for the meaning of the elements).

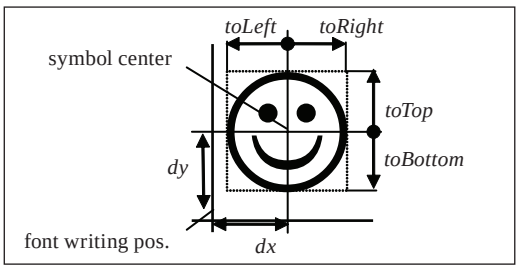
The rules for a group of symbols or for a specific symbol can be defined in the same way as for other symbols related to a note; for example, considering that the symbols in the group have to be positioned opposite to the stem in single voice and on the stem when in multivoice:

```
<symbolDirectionRule ruleId=”facesOppStem”>
        <group>faces</group>
        <oppositeToStem/>
</symbolDirectionRule>
<symbolDirectionRule ruleId=”facesOnStem”>
        <group>faces</group>
        <onStem/>
</symbolDirectionRule>

<applyRule rule=”facesOppStem”>
    <condition>
        <note>
            <inSingleVoice/>
        </note>
    </condition>
</applyRule>
<applyRule rule=”facesOnStem”>
        <condition>
            <note>
                <inMultivoice/>
            </note>
        </condition>
</applyRule>
```

As to the symbol position, other constraints can be specified: it has to be positioned on a space (not on a staff line) and outside the staff. This in both cases that the symbol is on the stem or opposite to stem:

Figure 20. meaning of dimension elements



```
<symbolPositionRule ruleId=”facesPos”>
    <group>faces</group>
    <onSpace/>
    <outsideStaff/>
    <dx>0</dx>
    <dy>0</dy>
</symbolPositionRule>

<applyRule rule=”facesPos”>
    <condition>
        <symbol>
            <oppositeToStem/>
        </symbol>
    </condition>
</applyRule>
<applyRule rule=”facesPos”>
    <condition>
        <symbol>
            <onStem/>
        </symbol>
    </condition>
</applyRule>
…
```

The symbol name (and not the group name) has to be used in the order rule to state the vertical relation of the symbol with other symbols, when it is placed on a note/chord:

```
<symbolOrderRule>
        <symbol>STAC</symbol>
        <symbol>TEN</symbol>
        <symbol>LEG</symbol>
        <symbol>MARTF</symbol>
…
        <symbol>PIZZICATO</symbol>
        <symbol>smile</symbol>
        <symbol>star5</symbol>
        <symbol>sad</symbol>
        <symbol>STRING</symbol>
…
</symbolOrderRule>
```

hence, the “smile” symbol has to be over the pizzicato and below the “star5” symbol.

SM-SI: SMR SYNCHRONIZATION INFORMATION

The SM-SI format contains information about the synchronization of the music score display with the whole multimedia scene where the music score is shown. For example, a multimedia scene may contain a video of an opera to be displayed synchronously with the music score of the same opera. Another possible case could be that the opera is a live event streamed over the network or through satellite and the music score has to be shown in synchronization with the live event.

The SM-SI is a binary format used inside MPEG-4 access units to provide to the SMR decoder the information needed to manage the synchronization. It has a different representation when it comes to synchronization with an off-line scene and with an online scene. In case of an off-line scene, the SM-SI data contain the sequence of measure numbers to be shown, with their corresponding duration expressed in milliseconds. Please note that the same measure can be repeated (with different durations) as it occurs with refrains. In the case of an online scene, the access unit coming from the transport layer (e.g., the network) with the SM-SI content contains the measure number to be displayed.

Figure 21. Notes with user defined symbols

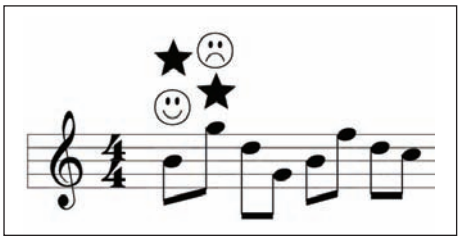
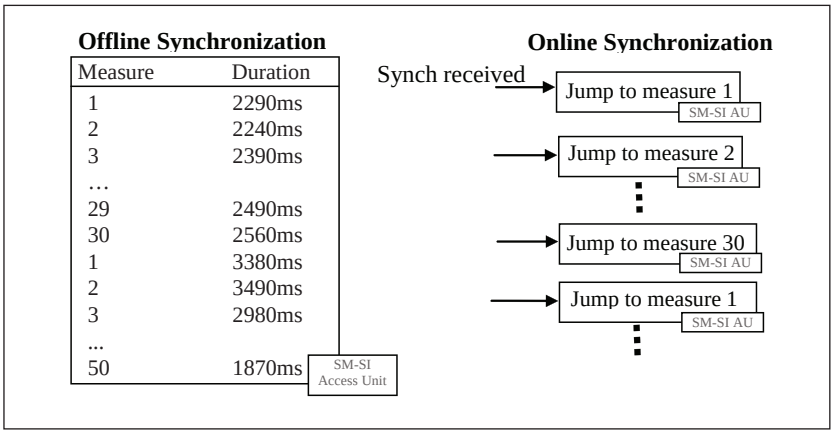


Figure 22. Off-line and online synchronisation using MPEG-SMR SM-SI



AUDIO GENERATION AND USER INTERACTION

It is possible to generate a MIDI representation of the music score by using the information contained in the score, namely, the metronome indication, the clef, and key signature currently active, while considering the dynamic indications included in the score. For transposed instruments, it is possible to indicate the transposition that should be used when generating the audio description. Moreover, each note can have both a visual indication of the note position on the staff and the note pitch. Whenever one of the two is missing, the other is calculated using the contextual information (clef and key signature). If the two elements are in contrast, the pitch is used for audio generation and the note position is used for rendering. Some parameters control the volume of dynamic indications (e.g., "f," "p," "mp," etc.), the MIDI instrument to be used for the score, the volume to be used for the instrument, and so forth.

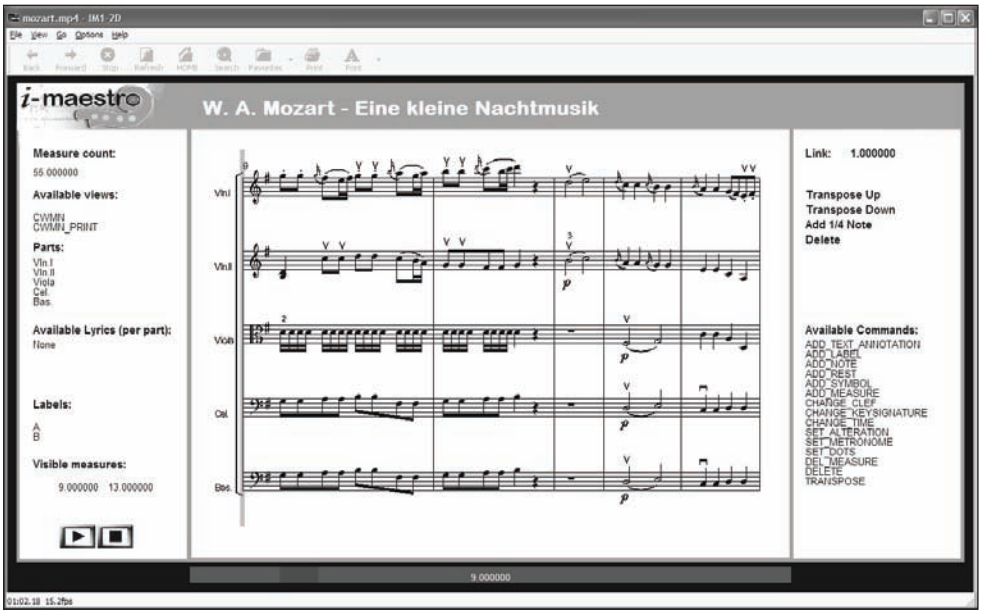
The MIDI generated may be used as a score language by the structured audio engine, and together with the orchestra language (defining how to produce the audio), it can produce the synthetic audio.

The SMR format encoded as MPEG-4 files can be used by any application and not necessarily restricted to MPEG-4 conformant players only. For these applications, no specific requirements for user interaction are present. However, an MPEG-4 conformant player that supports BIFS scene representation (a binarization of a VRML extension) can use the MusicScore node (defined in ISO/IEC 14496-11 AMD 5) to display the music score, coming from an SMR stream, into the multimedia scene. The player can display the score synchronously with other media (natural audio, natural video, or 3-D models), and allows some interaction with the SMR content.

The MusicScore node manages the rendering and display of the score, and supports the following functionalities:

- Browse the score (jump to a label, jump to a measure, jump to a given time).
- Choose which parts/views to see (single parts, main score, main score with only some parts).
- Choose the lyrics to be displayed (in case of lyrics in different languages).
- Generate an event when a music element with an associated link is selected.
- Start/stop play the score in synch with the multimedia scene, highlighting the current playing position in the score.
- Transpose the score.

Figure 23. IM1 MPEG-4 Player showing a music score



- Perform some editing of the score (e.g., add textual annotation, add/remove notes, add/remove symbols, …).

Using the MusicScore node with the other nodes managing user interaction, like the TouchSensor node, the MPEG-4 Player can realize an interactive scene where the user can use and manipulate the SMR content. Figure 23 presents a snapshot of the IM1 MPEG-4 player (part of the MPEG-4 reference software) showing the MusicScore node.

CONCLUSION

This chapter presented the MPEG-SMR music representation formats that are under standardization in the ISO/IEC MPEG group. The SM-XF, SM-FL, and the SM-SI languages have been discussed in some details with a number of examples. SM-XF XML language is used to model the music notation as main score, single parts, and lyrics, while SM-FL is used to represent the formatting rules of SMR, and SM-SI binary format is used to transport synchronisation information with the multimedia scene. Other aspects of the integration with the other MPEG tools, like BIFS, are treated in another chapter of this book.

The presence of MPEG-SMR will enable and support the realisation of many new applications in the areas of education, entertainment, and cultural valorisation. Currently, most of these applications are not available on devices such as I-TV, mobiles, and so forth, and applications available on PC are not based on standard content formats. The absence of such a standard is constraining producers to create their own formats. This is one of the key challenges for the diffusion of music knowledge and for the market of music education.

It is hoped that SMR will become a way for multimedia frameworks, and particularly for MPEG to express a great potential in the domain of music enjoyment and fruition, and also in all its related usages and practices, including music education and musical performance.

Further information on MPEG SMR can be found via the MPEG ad-hoc group on SMR Web

page, http://www.interactivemusicnetwork.org/mpeg-ahg, where a large collection of documents reporting requirements, scenarios, examples, and links are available.

ACKNOWLEDGMENT

Thanks to Paolo Nesi and Giorgio Zoia for the help provided, and thanks to everybody who has participated in the discussions on the reflector of the MPEG ad-hoc group on SMR, all members of the MUSICNETWORK WG on Music Notation, and everyone who attended joined meetings of MPEG and MUSICNETWORK. Special thanks (in no particular order) to Jerome Barthelemy, Kia Ng, Tom White, James Ingram, Martin Ross, Eleanor Selfridge Field, Neil McKenzie, David Crombie, Hartmut Ring, Tillmann Weyde, Jacques Steyn, Steve Newcomb, Perry Roland, Matthew Dovey, and many, many others (apologies for not being able to mention everyone involved here).

Part of the development on SMR has been performed under the i-Maestro EC IST project cosupported by the European Commission under the 6th Framework Programme.

REFERENCES

Bellini, P., Barthelemy, J., Bruno, I., Nesi, P., & Spinu, M. B. (2003). Multimedia music sharing among mediateques: Archives and distribution to their attendees. *Journal on Applied Artificial Intelligence, 17*(8-9), 773-795.

Bellini, P., Bruno, I., & Nesi, P. (2004). Multilingual lyric modeling and management. In S. E. George (Ed.), *Visual perception of music notation: On-line and off-line recognition*. Hershey, PA: IRM Press.

Bellini, P., Bruno, I., & Nesi, P. (2005). Automatic formatting of music sheets through MILLA rule-based language and engine. *Journal of New Music Research, 34*(3), 237-257.

Bellini, P., Della Santa, R., & Nesi, P. (2001, November 23-24). Automatic formatting of music sheet. In *Proceedings of the 1st International Conference on WEB Delivering of Music* (pp. 170-177). Florence, Italy: IEEE Press.

Bellini, P., Fioravanti, F., & Nesi, P. (1999). Managing music in orchestras. *IEEE Computer*, September, 26-34. Retrieved from http://www.dsi.unifi.it/~moods/

Bellini, P., & Nesi, P. (2001). WEDELMUSIC FORMAT: An XML music notation format for emerging applications. In *Proceedings of the 1st International Conference of Web Delivering of Music* (pp. 79-86). Florence, Italy: IEEE press.

Bellini, P., & Nesi, P. (2004). Modeling music notation in the Internet multimedia age. In S. E. George (Ed.), *Visual perception of music notation: On-line and off-line recognition*. Hershey, PA: IRM Press.

Bellini, P., Nesi, P., & Spinu, M. B. (2002). Cooperative visual manipulation of music notation. *ACM Transactions on Computer-Human Interaction, 9*(3), 194-237.

Bellini, P., Nesi, P., & Zoia, G. (2005). Symbolic music representation in MPEG for new multimedia applications. *IEEE Multimedia, 12*(4), 42-49.

Blostein, D. & Haken, L. (1991). Justification of printed music. *Communications of the ACM, 34*(3), 88-99.

Byrd, D. A. (1984). Music notation by computer. (Doctoral Dissertation, Indiana University). *UMI, Dissertation Service*. Retrieved from http://umi.com

CANTATE project. (1994). Deliverable 3.3: Report on SMDL evaluation, *WP3*. Retrieved from http://projects.fnb.nl

Capella. (2005). *CAPXML*. Retrieved from http://www.whc.de/capella.cfm

CUIDADO: *Processing of music and Mpeg7.* Retrieved from http://www.ircam.fr/cuidad/

Good, M. (2001). MusicXML for notation and analysis. In W. B. Hewlett & E. Selfridge-Field (Eds.), *The virtual score representation, retrieval, restoration* (pp. 113-124). Cambridge, MA: The MIT Press.

IMUTUS project. Retrieved from http://www.exodus.gr/imutus/

MOODS project. Retrieved from http://www.dsi.unifi.it/~moods

MPEG SMR AHG Web page. Retrieved from http://www.interactivemusicnetwork.org/mpeg-ahg

NIFF Consortium. (1995). *NIFF 6a: Notation interchange file format.*

Pereira, F., & Ebrahimi, T. (Eds.) (2002). *The MPEG-4 book*. Los Angeles, CA: IMSC Press.

Rader, G. M. (1996). Creating printed music automatically. *IEEE Computer*, June, 61-68.

Selfridge-Field, E. (Ed.) (1997). *Beyond MIDI—The handbook of musical codes.* London: The MIT Press.

SMDL ISO/IEC. (1995). *Standard music description language*. ISO/IEC DIS 10743.

This work was previously published in Interactive Multimedia Music Technologies, edited by K. Ng and P. Nesi, pp. 26-49, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.9
Possibilities, Limitations, and the Future of Audiovisual Content Protection

Martin Schmucker
Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany

ABSTRACT

This chapter explains the fundamental principles of audiovisual content protection. It explains the basic knowledge that is needed to understand the fundamentals of digital rights management (DRM) systems and their problems. Starting with a general introduction about copyright and content protection, available protection technologies are described and analyzed. The basic concepts of DRM solutions are explained and problems discussed. Potentials and practical limitations are analysed based on the digital music industry value chain. An outlook is given on solutions that are under development and that stronger consider the needs of the customers. In the conclusion, future solutions are discussed.

INTRODUCTION

I am about to put forward some major ideas; they will be heard and pondered. If not all of them please, surely a few will; in some sort, then, I shall have contributed to the progress of our age, and shall be content.

Marquis de Sade

Social and technical progress is one of the key issues of mankind. It is driven by the desire to disburden and to beautify life. Technical progress can be perceived in tangible goods like new devices, tools, and machines, while social progress cannot be perceived as easily. Both are based on a creative process resulting in new inventions and new ideas.

In law, the importance of this creative process is reflected by intellectual property (IP). "The term intellectual property reflects the idea that

this subject matter is the product of the mind or the intellect" as explained in "Intellectual property" (Wikipedia). Furthermore, it is explained, that IP rights (IPR) are more complex in the digital domain.

As the aim of IPR protection is to encourage new inventions, inventors, as well as authors, are granted (time-limited) exclusive rights for the exploitation of their works. Wikipedia identifies different intangible subject matters that are protected by IP laws:

- Copyright
- Patent
- Trademark
- Industrial design right
- Trade secret

As this chapter deals with digital (audiovisual) content, its focus is on copyright. The reader, however, should be aware that the described technologies are protected by patents and the mentioned products are protected by trademarks. This chapter is an updated summary of the technical report by Schmucker (2005c).

Copyright's main purpose is to prevent people from copying a person's original work. Copyright lasts for a specific duration.[1] After this well-defined period of time, the work enters public domain. The term copyright is generally equivalent to author's rights. Although certain organizations, like World Intellectual Property Organisation (WIPO), prefer the term author's rights, copyright is used within the area of DRM. The United States Copyright Office provides some general information on copyright and related issues (http://www.copyright.gov/).

Cohen (2002) describes how copyright changed due to the appearance of online work: Initially, copyright did not control access to, or private use of, an already purchased copy. Neither did copyright interfere with fair use derivatives. Nowadays, content owners claim the rights to control the access to, and the use of, content.

Music Copyright

Music copyright is a negative right, which means it gives the composer the right to restrict others from certain activities including copying music. Third parties who do not acknowledge these restrictions are liable for copyright infringements. Copyright automatically arises upon the creation of content without any formal registration process. Thus, copyright is distinct from other subsequent copyrights.

Music copyright includes different exclusive rights. As described in detail by Bell (2007), according to the U.S. copyright, a copyright owner has the right to:

- Reproduce copyrighted work
- Prepare derivative works
- Distribute copies
- Perform the work publicly
- Perform sound recordings of the work publicly

Further information on (music) copyright can be found, at:

- World Intellectual Property Organization (http://www.wipo.int/)
- Euro-Copyrights.org (http://www.euro-copyrights.org/)
- United States Copyright Office (http://www.copyright.gov/)
- Copyright for music librarians (http://www.lib.jmu.edu/org/mla/)
- "Copyright Issues for Music" at University of Buffalo Libraries
- 10 Big Myths about copyright explained by Templeton (2007)

Publishing Rights and Licensing

Copyright owners have the exclusive right to reproduce or make copies of their work. The copyright owner also has the exclusive right to

perform publicly a copyrighted work, directly or indirectly through means of communication or transmissions.

While these two rights (recording and public performance rights) were clearly separated before the digital distribution of content via the Internet, today this is not so clear anymore. Hence, service providers are forced to obtain multiple licenses from different parties. This process can be very difficult, as typically each right is connected with certain limitations, for example, geographical restrictions that can hardly be verified in the Internet. Rights can be negotiated either with the rights owner or with collection societies.

Legislation

International agreements that protect artistic and literary works aim to harmonize legal definitions and terms of protection. In the Berne convention, which was signed by more than 1,979 member states in 1979, such an international framework was determined. Yet, this framework has a degree of freedom to deviate from: As described by Cohen (2002), the copyright industries had secured an international commitment to additional legal protection for technological protection regimes in the 1996 WIPO Copyright Treaty, which leaves member states substantial flexibility in implementation.

The Digital Millennium Copyright Act of 1998 (DMCA, U.S.) forbids circumvention of access control technologies, and also the manufacture and distribution of circumvention devices. Hence, usage controls are protected indirectly. One side effect of the DMCA is shown by Craver et al. (Craver, Wu, Liu, Stubblefield, Swartzlander, Wallach, Dean & Felten, 2001) in the SDMI hack where the content industry tried to stop Felten distributing his research knowledge, cf. Felten (2001).

In Europe, the digital copyright directive was approved. This results in the classification of a range of devices that are to be prohibited. Yet, it leaves member states free to define what constitutes adequate legal protection against the act of circumvention. Member states may require preservation of exceptions (e.g., private noncommercial copying)

Furthermore, there are other legal frameworks, for example, the Uniform Computer Information Transactions Act (UCITA), which would validate consume "assent" to these restrictions, and legitimise the accompanying technological controls as part and parcel of the agreement.

Other potential areas of conflict are fair use and privacy. The possibilities of DRM are usage control. A qualification of the usage, however, is difficult. For example, devices and computers cannot distinguish between legal and illegal copying. Thus, a restrictive policy is enforced by DRM: No copying is allowed.

This led to interesting court decisions like in France, where the French court ruled that copy protection schemes have to be removed from DVDs, as described by Orlowski (2005). Similarly, the Deutsche Bibliothek (now German National Library) signed an agreement with the content industry that allows one to crack and to duplicate DRM-protected digital media. Here, DRM opposes the legal mandate of the German National Library.

The storage and exchange of personal information is sometimes considered critical as well. For example, a customer exchanges information with a third trusted party, which stores this information. This potentially infringes privacy of the customers, which is analysed, for example, by Grimm (2005).

Illegal File Sharing

Obviously, for very good reasons, content sharing is restricted. In the digital world, IP theft is not as obvious as theft in the physical world. In contrast to physical goods, digital content can be easily reproduced without quality loss. The copy is the original. Thus no other person experiences a lack of the digital content that was stolen.

Nevertheless, file sharing is against the copyright unless:

- The content is in public domain.
- The owner/creator gave permission to share it.
- The content is available under prosharing license (e.g., Creative Commons, http://creativecommons.org/).

As a result, the content industry successfully identified, and still identifies, users who illegally distribute digital content on the Internet via Web sites or via Peer-to-Peer (P2P) networks. Each user can be identified through the corresponding unique IP-address. This is required to receive information from the Internet. For permanent IP addresses, this is very easy (e.g., in the case of companies). In the case of dynamic IP-addresses (dial-in access), the Internet Service Providers (ISPs) temporarily store this information. The content industry was successful in identifying users by requesting user-related data from the ISPs.[2]

Nevertheless, this procedure is under strong discussion, and different verdicts exist in the different countries. Also the resulting lawsuits and their success are heavily discussed. Rudish (2005) states that "Eight of nine lawsuits filed last summer against Emory students accusing them of illegally sharing copyrighted music files have been dismissed, according to Senior Vice President and General Counsel Kent Alexander. The Recording Industry Association of America (RIAA) spokesperson Jenni Engebretsen said that one of the Emory cases has been settled, but she could not confirm the dismissal of the others."

Digital Rights Management

The term digital rights management (DRM) comprises technologies that allow the usage control of digital content. This goes beyond the possibilities copyright holders had before, thus DRM threatens user privileges. For example, fair use[3] or archiving of content is also restricted, as DRM systems restrict access to content. Unfortunately, there is no unique definition for DRM. Thus, when somebody faces the term DRM, the connotation associated with it must not be neglected.

In "Digital Rights Management and Libraries," a selection of different connotations is given:

- "Digital rights management technologies are aimed at increasing the kinds and/or scope of control that rights-holders can assert over their intellectual property assets." (as taken from *Electronic Frontier Foundation*, http://www.eff.org/)
- "DRM must be about the 'digital management of rights' not the "management of digital rights." (as described by the *W3C Workshop Report on DRM for the Web*, http://www.w3.org/2000/12/drm-ws/)
- "The purpose of DRM technology is to control access to, track and limit uses of digital works." (as seen by *The American Library Association*, http://www.ala.org/, itself)
- "DRM are the technologies, tools and processes that protect intellectual property during digital content commerce..." (as defined by the *Publishers' Requirements for DRM, W3C Workshop Report on DRM for the Web*, http://www.w3.org/2000/12/drm-ws/minutes/publishers.html)
- "DRM systems restrict the use of digital files in order to protect the interests of copyright holders." (as taken from *Electronic Privacy Information Center*, http://www.epic.org/)

Camp (2003) outlined the copyright system's legal, technological, and economic foundations with the aim to support the design of DRM systems. She identified several key functions[4], which should be considered in the requirements of a DRM system. Among these key functions identified by Camp are:

- Protection of the author's reputation
- Protection of the work's monetary value
- Archiving of content
- Ensuring of content integrity
- Providing surety through persistence[5]
- Facilitating personalization through filtering and annotation[6]

Although copyright defines under which circumstances copying is legal and when copying is illegal, copyright infringements are ubiquitous. Therefore several campaigns were launched addressing this topic. As discussed by Walter (2003), the MPAA launched an advertising campaign, copying is stealing, to sensitive public that IPR infringement by private people can be compared to stealing a CD from a record shop.

Similar threats to music can be identified even before the predigital age (before the 1980s), for example, shortly after the completion of the first pianolas in 1895 (http://www.pianola.org/), a Pianola copyright ruling was cited in 1899, according to Rhodes: "Boosey vs Whight (1899) involved copyright charges arising over the production of pianola rolls, in which the court found that the reproduction of the perforated pianola rolls did not infringe the English copyright act protecting sheets of music."

When analogue audio tapes and also video tapes emerged, potential threats caused by illegal copies were realized. As by Walter (2003), in the predigital age, several legal disputes are known where copyright owners claimed copyright infringement offences:

- Ames Records allowed subscribers to hire records from it for a small rental charge.
- Amstrad supplied tape-to-tape recording equipment.
- Sony's video recorders were used for illegal copying.

Interestingly, neither Ames nor Amstrad nor Sony was liable for copyright infringement. Before the introduction of the compact disk (CD) in 1982, music was typically sold on long playing (LP) vinyl records. The sales of LPs slowly declined with the introduction of the music cassette (MC), which allowed copying of music, cf. Lewis (2003). Content was, however, stored in an analogue representation. Copying this analogue content was not possible without loss of quality. Therefore a natural barrier existed limiting the amount of recopies. In addition to these natural barriers, copy prevention systems were developed. This natural barrier no longer exists in the digital world.

Nowadays, copying digital data is much easier, and commercially oriented pirates, as well as some consumers certainly do misuse this: Digital data can be copied without any loss of information, and distributed fast world wide via the Internet. Especially P2P file-sharing networks—the most popular one was probably Napster (http://www.napster.com/)—enable users to share content. After the US courts shut down the first version of Napster[7], rights holders still claim that Napster's descendants cost billions of dollars in revenues.

While the rights holders were quite successful against Napster, actions against other P2P-software suppliers like Grokster (http://www.grokster.com) and StreamCast (http://www.streamcastnetworks.com/) failed. These service suppliers cannot control the use of the technology by the end user, and the users' communication is entirely outside the control of the service suppliers. Today's descendants have a decentralized architecture and cannot be shutdown easily. As a consequence, rights holders now target consumers, ISPs, operators, and even founders of file sharing systems.

The content industry, especially the music and movie industry, nowadays sues P2P users who exchange content illegally. At the beginning of these activities, the P2P users reduced their illegal file exchange as outlined by Greek (2004). This resulted from the news that P2P users are sued for their IPR infringements. Additionally, the content industry started PR campaigns to raise

the users' awareness for the illegality of the file sharing of copyright protected content. Despite these activities, illegal P2P-usage seems to have increased again, and users have identified other ways of exchanging content, as discussed in Madden and Rainie (2005): "Beyond MP3 players, email and instant messaging, these alternative sources include music and movie Web sites, blogs and online review sites." Also, with portable storage devices, sharing content is very easy and convenient. Recent news related to P2P systems is available, for example, on People to People net (http://p2pnet.net/).

Although Verizon RIAA won a court order forcing an Internet Service Provider to disclose the identity of individual consumers who traded music files, technologies like Freenet[8] (http://freenet.sourceforge.net/) allow users to share any kind of content strongly reduced risk of being identified by rights holders. Obviously, P2P developments reacted on the content industries activities: While the first generation of P2P-file sharing networks has a centralised file list, the second generation is a purely distributed architecture. And the third generation addresses the anonymity of its users as outlined in "Peer-to-Peer" (Wikipedia).

Business aspects cannot be neglected when discussing and analysing protection technologies and illegal distribution. Unfortunately, there is no unique view on the influence of P2P exchange to the development of the traditional and the online market. On the one hand, each copied file is considered as a loss. On the other hand, some people consider downloaded content as an appetizer. Different reports and studies exist where common people have been interrogated about the influence of P2P networks. The results of the studies are contradicting, comparable to the results of studies trying to identify the reasons for the decrease in CD sales.

Besides these previously discussed methods and procedure, the technical endeavours of controlling the usage of content are summarized in the term digital rights management (DRM), and were first focused on security and encryption addressing the problem of unauthorized copying. But DRM evolved, and now it covers various issues including:

- The description of content
- The identification of content
- Trading and exchanging content
- Protection of content
- Monitoring/tracking of content distribution and its usage

As emphasised by Iannella (2001), DRM is the "digital management of rights" and not the "management of digital rights." Thus, it has become a very complex area addressing issues far beyond security and encryption.

The first section gives an overview of the current situation of copyright and content distribution. Section two introduces the available technologies for active and passive content protection. The following section, three, describes the application of individual techniques in DRM solutions. Their possibilities and limitations are discussed in the next section. New developments that try to embrace user requirements are described in the fifth section. An outlook discusses the future development of content protection of digital audiovisual content.

AVAILABLE TECHNOLOGY FOR AUDIOVISUAL CONTENT PROTECTION

Engineers like to solve problems. If there are no problems handily available, they will create their own problems.
Scott Adams

In this section, the basic technologies are described that are available for content protection. Some of these technologies are also relevant for other areas, like the content identification and the linkage of content and metadata.

Content Identification, Content Description, and Content Management

As already outlined before, content identification is an important aspect when content-related information has to be identified. For example, this is a central aspect of libraries and archives, where content-related metadata is managed.

Whenever data has to be accessed or retrieved, two issues are important:

- Content identification
- Content description

These issues are independent of DRM. But also DRM has to somehow identify content, as usage and rights information relate to specific pieces of content. In the following, a very brief overview on different standards for content identification and descriptions are given.

Content Identification

Content identification should be accomplished with an open standardized mechanism. Several open standards have been created for this purpose in the digital world. They allow identifying content or resources uniquely. Among the most commonly used are:

- International Standard Book Number (ISBN, cf. http://isbn-international.org/ or http://www.isbn.org/)
- International Standard Serial Number (ISSN, cf. http://www.issn.org)
- International Standard Music Number (ISMN, cf. "International Standard Music Number")
- Uniform Resource Identifier (URI, cf. "RFC1736", "RFC1737" and "RFC2396")
- Digital Object Identifier (DOI, cf. http://www.doi.org/)
- International Standard Text Code (ISTC, cf. "International Standard Text Code")

Content Description

Within a DRM system itself, the most significant content description is the licensing information. For completeness, general content description is summarized briefly.

According to Iannella (2001), content description should be based on the most appropriate metadata standard for each genre. Any overlap with other metadata systems might result in difficulties in the implementation due to redundant information.

Among the existing standards for content description are:

- Online Information Exchange (ONIX) as developed by EDItEUR (http://www.editeur.org/).
- IMS Learning Resource Metadata Information Model by IMS (http://www.imsproject.org/)
- Dublin Core Metadata Initiative (DCMI, http://dublincore.org/)
- Interoperability of data in eCommerce systems <indecs> (http://www.indecs.org/)
- "EBU metadata specifications" from the European Broadcasting Union
- Standard Media Exchange Format (SMEF), cf. "SMEF Data Model"
- MPEG-4 defines a stream management framework. This framework includes a rudimentary representation of metadata for the "description, identification and logical dependencies of the elementary streams" (http://www.chiariglione.org/mpeg/).
- MPEG-7 addresses the describing of and searching for content. "MPEG-7, formally named "*Multimedia Content Description Interface*," is a standard for describing the multimedia content data that supports some degree of interpretation of the information's

meaning, which can be passed onto, or accessed by, a device or a computer code." (http://www.chiariglione.org/mpeg/).

Rights Management and Rights Description Languages

As DRM is the digital management of rights, they have to be represented in a digital format to be digitally manageable. These digital representations must consider several aspects, as also described by Rosenblatt et al. (Rosenblatt, Trippe, & Mooney, 2002):

- **Content rights transactions:** Traditional business models.
- **Components of rights models:** Types of rights and their attributes.
- **Fundamental rights:** Render rights (print, view, play), transport rights (copy, move, loan), derivative work rights (extract, edit, embed).
- **Rights attributes:** (considerations, extends, types of users)

A general problem of DRM systems is the fact that they are not able to qualitatively distinguish between the different kinds of usage. For example, copying for personal purpose and copying for friends or even unknown persons is represented as the same action within a DRM system. This is what Rosenblatt et al. (2002) expressed as "they [digital rights models] don't do a great job of modeling the actual uses of content."

This is a potential starting point for further developments. The complexity of such an approach, however, might be beyond the relatively simple license (models), not only due to the potential dynamics of sociocultural aspects.

Today's licenses can have a strongly varying range of usage rights and conditions reflecting everything from simple to complex rights situations. Therefore, the language used for the description of rights should be able to model even very complex situations, which can appear easily when dealing with digital content (e.g., audiovisual material).

Rights Description Languages

The purpose of a digital license is to express who can do what with a specific content under certain conditions. For the digital management of rights, obviously, this license has to be expressed in a machine-readable way.

The extensible markup language (XML) is a de facto standard for the exchange and storage of data. Due to its flexibility, several rights description languages are based on XML. An overview of different XML-based rights description languages can be found in "XML and Digital Rights Management (DRM)." The currently most relevant ones are:

- Digital Property Rights Language (DPRL) and its successor eXtensible Rights Management Language (XRML, http://www.xrml.org/) are implemented by ContentGuard (http://www.contentguard.com/) and became an official standard within MPEG-21 (ISO/IEC 21000).
- Open Digital Rights Language (ODRL, http://www.odrl.net) was adopted by the Open Mobile Alliance (OMA).

Rights Processing

Expressing the rights in a machine-readable way is only the first step. Ideally the rights are processed automatically whenever content is created, derived, or exchange. This cannot be achieved yet as different terminology, especially when dealing with multinational content, and even different legal foundations complicate an automation process. Thus a rights ontology or thesaurus is inevitable.

Current Status of REL Standardisation

MPEG so far chose XrML as a basis for the MPEG Rights Expression Language (REL). Nevertheless, OMA was in favor of ODRL. This competition between XrML and OMA is very interesting, as XrML is patented and ODRL is royalty free. Nevertheless, "ContentGuard asserts that because its patents cover DRM implementations based on any rights language, even ODRL implementations should be subject to patent licensing from ContentGuard. The OMA tacitly disagreed with ContentGuard's assessment when it chose ODRL; the issue has yet to be tested, in the courts or otherwise" as pointed out by Rosenblatt (2004).

MPEG REL seems to be too complex and is hardly accepted by the market. Microsoft still prefers XrML. In March 2007, the first publicly known DRM patent licensing deal was done by ContentGuard. LG Electronics will apply XrML in it mobile handsets. A more recent overview of DRM related activities can be found at DRMWatch (http://www.drmwatch.com).

Encryption

Whenever data is transmitted over an insecure channel, which indeed is the Internet, the only possible protection mechanism to guarantee confidentiality is encryption.[9] The methods used for encryption can be attacked. These attacks are not limited to the encryption algorithm itself. Attacks are also possible against keys or protocols. In this section, we will address some general aspects of encryption to allow a basic understanding of distribution systems and related requirements. Detailed information on encryption and cryptography was given, for example, by Menezes et al. (Memezes, Oorschot, & van, Vanstone, 1996) or Schneier (1996).

Cryptography

Cryptography is the art of encryption and is several thousand years old. Encryption transforms the content by using an encryption algorithm or a cipher. Retransformation of the original message (or plain text) from the encrypted form (or cipher text) is known as decryption. To prevent others from reading the cipher text, the method could kept secret or the algorithm uses a secret to determine the transformation. Kerkhoff (1883) already formulated in 1883 that security by obscurity is not possible: Keeping the encryption method secret does not increase the security of the method. The security of an algorithm therefore must not be based on its secrecy but on the usage of a key.

Different methods exist:

- Symmetric encryption methods
- Asymmetric encryption methods

Symmetric Encryption Methods

Whenever data is exchanged, communication partners agree on a common key for the encryption of the data, as shown in Figure 1. As the same key is used for encryption and decryption by symmetric encryption methods, everybody who has access to the key can decrypt encrypted data.

Symmetric encryption algorithms include:

- **Substitution algorithm:** By using a table every character is replaced by another one.
- **Vignere method:** A password determines the mapping.

One severe attack is the knowledge of an instance of encrypted and the decrypted data, that allows the calculation of the mapping and therefore, the decryption of other encrypted messages. But also statistical attacks can be applied. Of course the system security depends on the

Figure 1. Symmetric encryption methods use the same key for encryption and decryption. The key determines the transformation for the plain text to the cipher text. Thus, everybody who has knowledge about the secret key can decrypt the cipher texts, which have been encrypted with this key.

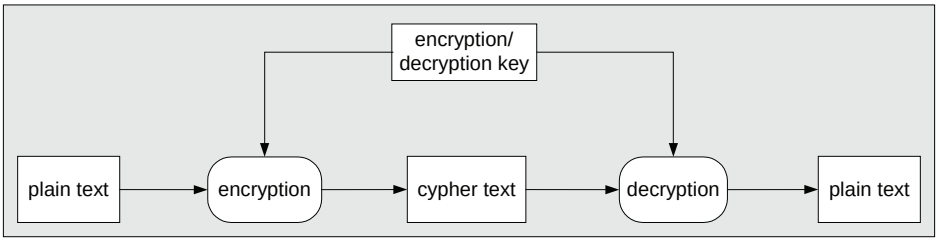
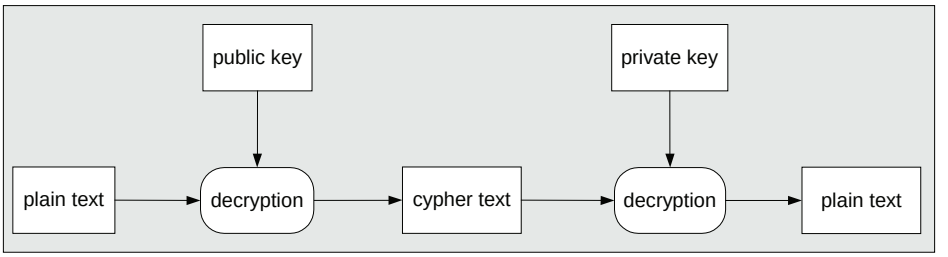


Figure 2. In contrast to the symmetric encryption algorithm, the asymmetric encryption methods use different keys for encryption and decryption. The public key can be accessed by everybody interested in encrypting a message for a certain receiver. The private key is kept secret.



length of the password or key. If the key has the same length as the message maximum, security is achieved. In this case the key is also called a one-time pad.

There are two main types of symmetric encryption algorithms that differ in the size of the data the cipher works on:[10]

- Block ciphers process a number of bits simultaneously, for example DES, 3DES, IDEA, AES, Blowfish, Twofish, and RC6.
- Stream ciphers process a single bit at a time, for example, RC4.

Asymmetric Encryption Methods

The general problem of the symmetric schemes is the exchange of secret keys. As the secret key has to remain secret, the transmission of keys in plain text is not possible. This problem is addressed by public-key or asymmetric encryption methods.

Asymmetric encryption methods use two keys:

1. The public key is for the encryption of data. This key can be distributed freely.
2. The private key is used for the decryption of the data.

Thus, no keys have to be exchanged. One can even think of a "telephone book" that publishes the e-mail addresses and the corresponding public keys. Yet, public key encryption is computationally expensive.

The existing solutions for the asymmetric methods are based on the computation of mathematical calculations that are extremely difficult for very large numbers, like done, for example, in:

- ElGamal
- RSA
- Elliptic Curve Ciphers (ECC) probably will replace RSA in the future

Key Length

The comparison of different ciphers based on their key lengths is meaning less: The security is influenced by the design of the ciphers. One example is public key algorithms' key length: They require much longer key lengths than symmetric algorithms.

Recommendations on the key lengths are given by different national organizations. A collection can be found at http://www.keylength.com/. These recommendations consider that advance of cryptoanalysis.

Cryptoanalysis

Cryptoanalysis deals with the analysis of cryptographic methods. For example, the "brute force" attack is a straight forward attack that calculates and verifies all possible keys. Of course this can be very time-consuming, but for certain encryption algorithms, hardware was developed to speed up this task, and even distributed calculations that use a huge amount of computers connected via the Internet are performed. A method can be considered as secure when the most effective attack is the "brute force" attack. However, cryptoanalysis is not limited to the decryption of the secret message: collecting any kind of information, which provides more information about the secret message, represents an attack.

Dangers and Attacks

The security of all asymmetric encryption methods depends on the computational complexity of the corresponding mathematical problems. A tricky calculation or quantum computers might endanger the security of all asymmetric encryption methods in the future.

Besides this potential risk, attackers can exploit other leaks like the previously mentioned randomness of a PRN generator. Further possible leaks are cryptographic protocols, chosen keys, short pass phrases, and so forth. Even more sophisticated attacks are applicable, like the ones based on the power consumption or the time delay of cryptographic coprocessors.

One-Way Encryption

Encryption with one-way algorithms[11] cannot be reversed. Typical applications are scenarios, where the plain text must not be recovered, and include the storage of passwords. One of those one-way hash algorithms is the secure hashing algorithm (SHA) that creates a 160-bit hash value. As these one-way encryption functions typically base their calculations on a password, they can be used to sign data with a digital signature.

Depending on the application, for example, like for authentication/verification of digital content, the security of one-way encryption is very significant. Recent attacks on MD5, as shown by Wang et al. (Wang, Feng, Lai, & Yu, 2004) or on SHA-1 MD5, as outlined by Wang et al. (Wang, Yin, & Yu, 2005), show that collisions are possible, and that MD5 and SHA-1 can no longer be considered secure.

Applications in DRM Systems

Encryption technologies are primarily used to secure the communication between different parties, and the storage at the parties' storage media. While this is reasonable in business environments, concerns have to be raised with respect to the encrypted storage of content at the consumers' side. As consumers access the encrypted data, an unencrypted version must be temporarily available in the memory of the computer. As a matter of fact, consumers that are capable of handling debugging software are also able to access this decrypted content as long

as trusted hardware solutions are not available. This is discussed next.

Besides the secure communication and storage of content encryption, technology is used for further applications:

- Verifying content based on digests.
- Verifying identities based on certificates.
- Verifying identities and content based on signatures.

Watermarking

Besides the active protection technologies, like the previously described encryption and cryptography, passive protection technologies provide further possibilities in protecting content. For example, passive protection technologies address the identification of content or the identification of content owners. Thus, they do not prevent copying, per-se. These mechanisms, nevertheless, can be used for the detection of IPR infringements, as shown in the different movie piracy cases where several Oscar screeners—among them were *The Last Samurai*, *Shattered Glass*, and *In America*—were illegally distributed on the Internet in illegal file-sharing networks (cf. "Arrest In 'Screener' Upload Case," "FBI Arrests Internet Movie Pirate," and "Man nabbed for uploading Oscar 'screener'").

This section describes digital watermarking techniques, which allow embedding arbitrary information directly into the any multimedia content imperceptibly. The embedded information depends on the application. For the protection of intellectual property, typically, information about the rights holder is embedded. Information about the content itself, or a link to corresponding metadata, can be embedded, which supports the identification of content. Other scenarios embed information relevant for authentication[12] or even information related to marketing and PR.[13] In some applications, for example, for transaction tracing, as it was in the case of the Oscar screeners, the embedded information might consist of a customer identifier.

On the one hand, perceptible watermarking techniques influence the quality of the multimedia content. On the other hand, a successful removal can be easily verified by an attacker. Therefore we will limit this discussion on imperceptible watermarking techniques.[14]

In contrast to steganography, where the most important requirement is that the communication remains undiscovered, in digital watermarking, the information, if a message was embedded or not, can be publicly available. This knowledge leads to increased requirements on the robustness and the security of the communication respective of the embedded message. Thus, a good watermarking system maximizes robustness for a constraint-perceived quality degradation.

Characteristics and Requirements

The general principle of watermarking methods can be compared with the symmetrical encryption.[15] Both methods require the same key for encoding and decoding the information. A watermarking system consists of two subsystems: the watermark writer (encoder) and the watermark reader (decoder).

The embedding process is shown in Figure. Watermarking technologies' most important characteristic is the active embedding of a watermark message (e.g., an identifier) in the content. Thus, the content is modified (imperceptibly). This message is read during the retrieval process, as shown in Figure 4. This has severe implications to the protection scenario as in some cases, unmarked content that is already distributed has to be considered carefully.

The retrieval processes (as shown in Figure 4) of existing methods differ in the detection itself as well as in the number of input parameters. Blind[16] detection schemes require only the marked content and the detection key for the detection. In contrast to the blind detection schemes, the

Figure 3. During the watermarking embedding process the watermarking message is interwoven with the original content

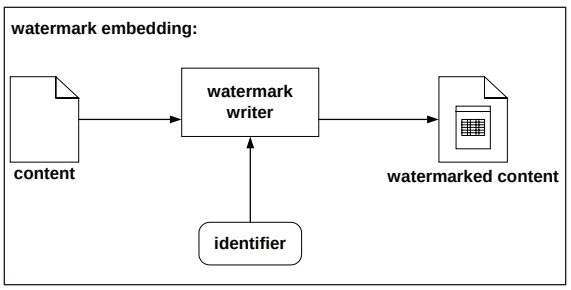
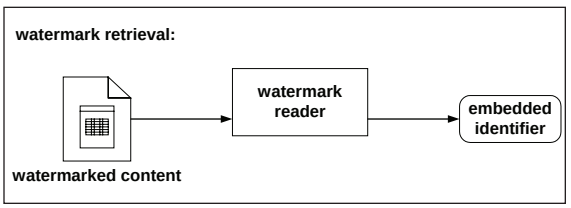


Figure 4. During the retrieval process the embedded watermark message is restored



nonblind[17] methods require, in addition to the previous parameters, the original and sometimes the embedded watermark message. Semiblind methods require, in addition to the blind detection schemes, the watermark message as a retrieval parameter.

Nonblind methods are practically relevant only for a very limited number of application scenarios due to the necessary input of the original. In the typical application scenarios, like broadcast monitoring or the automatic identification of the copyright owner, the original is either not known or not immediately accessible.

General requirements on watermarking techniques are:

- The quality can be considered as the most important criteria. In general, the embedding of a message into the content should not affect the perceived quality of the content (the information carrier). As perceived quality always depends on the media type, watermarking techniques have to be developed or adapted to individual media types (e.g., choosing a suitable perceptual model to minimize distortions).[18]
- The robustness is defined by the types and numbers of operations (and their parameters) applied to the watermarked content, which can be survived by the watermark message. From a watermark developer's and user's view, these processing operations are called attacks. Depending on the intention of the operations, they can be distinguished between intentional and unintentional attacks. Although an attacker has numerous attack operations available, their combination and their parameters cannot be chosen arbitrarily, as the result also has to fulfill a certain quality requirement as well. The operations a watermarking scheme should be robust against are defined by the individual application scenario. For the identification of content and protection of IPR robustness can be considered as the second most important criteria.
- The capacity is the amount of information, which can be embedded in the content. It is the third most important criteria. Due to the mutual dependencies between quality, robustness, and capacity, a certain quality level is defined (according to the application scenario), and the robustness is chosen dependent on the deserved quality. Capacity is finally defined by quality and robustness.
- The complexity of an algorithm is important for certain application scenarios where real-time embedding or detection is important.
- The security of a watermarking scheme does not only depend on the robustness, but also on other issues like the used key and the message embedding method. Also, its implementation and integration into the overall system cannot be neglected.

General information on watermarking techniques was collected by Katzenbeisser and Petitcolas (2000). Cox et al. (Cox, Miller, & Bloom, 2002) provide a detailed technical inside on watermarking schemes for images. An application-oriented introduction and detailed information about requirements and application scenarios is given by Arnold et al. (Arnold, Schmucker, & Wolthusen, 2003).

Limitations

Watermarking schemes are advantageous as they allow the embedding of arbitrary information directly into content. This embedded information can survive processing and media conversion. In the sense that compression techniques remove imperceptible information, they can be regarded as competitors. Thus, the effect of future emerging compression techniques on the embedded information is unclear.

In IPR protection scenarios, watermarking techniques certainly have some limitations. For example, the robustness of watermarking schemes might not be sufficient, and an attacker might be able to remove, or maybe copy, a watermark message. Nevertheless, an attacker cannot be 100% secure about the success of his attack.

Available objective tests and performance analysis of existing watermarking techniques address a limited scope. These benchmarking suites like Stirmark by Peticolas (2006) or Certimark (http://www.certimark.org) do not fully consider practical requirements. Here, standardized application scenarios defining requirements would be advantageous. Another limitation is the missing standardization of the embedded information.

Applications in DRM systems

Although watermarking techniques must be developed for individual media types, a broad range of watermarking algorithms are available for various media types including audio, images, video, geometry data, text, music scores, and so forth. Several requirements can be addressed by integrating watermarking techniques into DRM solutions, as also described by Rosenblatt (2002):

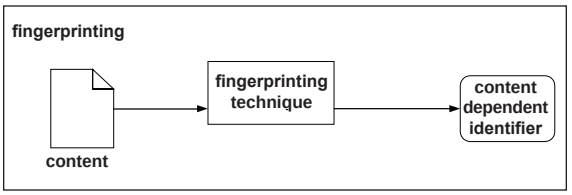
- Source identification can be achieved by imperceptible watermarking techniques, which do not affect the perceived quality. The information embedded links to the content owner or to a rights owner.
- Tracking and tracing by embedding so-called transaction watermarks, which is information about people involved in transactions, might allow the detection of leaks within the distribution chain.
- Metadata labelling stores a link to a database containing metadata information.

From a security point of view, applications involving encrypted watermarks and encrypted files with watermarks should be considered critically if they are used for access control in end-user devices. Yet, Rosenblatt et al. (2002) concludes "a scheme that incorporated both encryption and watermarking is not foolproof, but (all else being equal) it's the best DRM scheme available."

Fingerprinting or Perceptual Hashing

In contrast to watermarking techniques, which modify content, fingerprinting, or perceptual hashing techniques can identify content and fragments thereof without prior modifications. Thus, they have an inherent advantage if used in application scenarios where content cannot be modified, for example, as already distributed or due to workflow limitations. In this section, we explain shortly the idea and application of fingerprinting technologies.

Figure 5. The fingerprinting method calculates the content dependent identifier directly from the original content. Thus the content has not to be modified.



General Principle

Fingerprinting techniques calculate a content-dependent identifier, as shown in Figure 5. This content-dependent fingerprint can be compared with a human fingerprint: It is a unique identifier for renderings, and the original content cannot be created out of this identifier. Thus these techniques are also related to the cryptographic one-way functions. The significant difference to cryptographic hash functions is that cryptographic hash functions map similar input values not to similar hash values. For fingerprinting techniques, the opposite requirement must hold. Therefore, they are also called perceptual or soft hashing functions.[19] Perceptual hash reflects the fact that perceptual similar content should result in a similar hash value. Fingerprinting techniques are methods for content based identification (CBID).

Due to this property, fingerprinting solutions are very suitable for identification applications, like automatic play list generation or broadcast monitoring, as described by Allamanche et al. (Allamanche, Herre, Hellmuth, Froba, & Cremer, 2001). Further potential applications include the tracking of content flow or even restricting the content flow (e.g., in corporate networks). Due to these characteristics, fingerprinting techniques have attracted increased attention recently. A very good overview on the basis of audio fingerprinting is given by Cano et al. (Cano, Baltle, Kalker, & Haitsma, 2002).

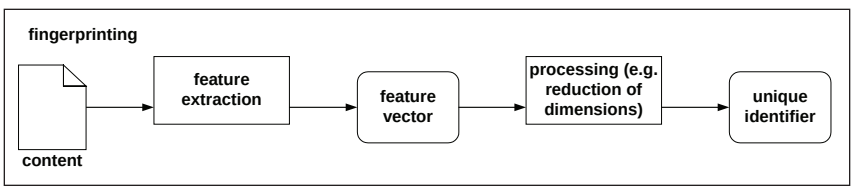
Characteristics and Requirements

Fingerprinting techniques are related to content-based retrieval (CBR). CBR methods create a content-dependent signature directly from the content. This signature is also compared to pre-calculated and stored content signatures.

Figure 6 shows the principle steps necessary for the calculation of a unique identifier. First, features are extracted from the content. These extracted features typically have very high dimensions and are processed further, resulting in unique identifiers.

The feature extraction process itself might extract features that are directly perceived by humans, like the frequency characteristics (melody) in the case of audio signals. On the contrary, features that do not directly depend on the human perceptions can also be used, as long as they allow the automatic discrimination of content. A typical example is the sign of the difference between neighboring (frequency) energy bands, as proposed by Haitsma et al. (Haitsma, van der Veen, Kalker, & Brueker, 2000).

Figure 6. For calculating a unique identifier, (perceptually) relevant features are extracted. These features are processed to reduce their dimensionality.



Without a following processing step, the dimensionality of the feature vector would be very high. Therefore, the dimensionality is reduced by removing redundant information. During this processing step, further improvements, for example, error resilience, can be achieved. Although the aim is to receive a compact digest, the discriminability of this digest still has to be sufficient.

In addition to the feature vector, a suitable distance measure is required. This is also related to CBR methods, where it defines similarity. Typical distance measures are the Euclidean or related distance measures or distance measures based on correlation.

The retrieval architecture strongly influences the complexity and the scalability of the fingerprinting technique. Efficient spatial access structures have been developed like indices, or application-oriented approach like the ones used for DNA information processing. In the retrieval architecture, a final hypothesis test calculates the probability of the correct identification of the content.

In the case of content identification, a good performance in discriminating a huge amount of data and corresponding fingerprints is crucial. Similarly to watermarking techniques, requirements can be identified:

- The robustness can be defined by the types and numbers of operations and their parameters applied the content, which does not effect the retrieval of the content. Typical operations depend on the application scenario where the fingerprinting method will be integrated in. For example, when a system should be able to recover the song from a radio transmission that is recorded via a mobile phone, the fingerprinting system should be capable of the reduced frequency band available due to the mobile phone. Also small audio extracts somewhere within the song must not result in misidentification, as humans will rather realize in the middle or at the end of the song that is worth being remembered or purchased. Finally, a noisy background will probably be the general recording place, for example, in a car, bar, club, or café.
- The discriminability determines the capability of how many content items can be identified.
- The scalability is an important practical criterion. Today, millions of different types of audio content exist. Some of them are even available in different editions, for example, studio or live performance recordings. In this case, a system should be capable of handling all available works in a reasonable amount of time, where "reasonable" is again defined by the application scenario.
- The complexity of an algorithm is important for certain application scenarios where real-time identification is important.

In contrast to watermarking applications, so far security is hardly considered. This might be caused due to the fact that perceptual hashing techniques, so far, are mainly applied in CBID scenarios, where security can be neglected. In authentication scenarios, however, security is crucial and requirements are different.

Limitations and Comparison to Watermarking

Different limitations have to be considered when a fingerprinting technology is deployed. As already mentioned, fingerprinting techniques do not modify the content, but calculate an identifier directly from the content itself. This is an obvious advantage when content is already available in a nonmarked version. Yet, this is also a drawback in comparison to watermarking schemes: personalization is not possible. Therefore, applications like leakage tracking are not possible. Although content can be tracked, tracking users is not possible, as content is generally not unique for individual users.

Instead of being marked, content must be registered. That means that only content can be identified if its fingerprint was previously calculated and stored in a database. And if the identifier is stored in a database, this database has to be accessible during the identification process.

Another limitation of fingerprinting techniques have to consider when using fingerprinting for controlling the data transfer networking infrastructure: encrypted content or scrambled content cannot be identified. Identification is only possible with content that is accessible as it is intended for rendering.

A comparison between the different properties of watermarking and fingerprinting is shown in Table 1.

Applications in DRM systems

Besides the previously listed applications of fingerprinting systems in automatic play list generation, broadcast monitoring, content tracking, and content flow limitations, another application is very interesting for fingerprinting techniques: royalty distribution. Content can be monitored in peer-to-peer networks with the help of fingerprinting techniques. This information can be combined with other information available, for example, metadata within the peer-to-peer networks used by humans for content identification. Keeping in mind the future revenue stream, new possibilities can be created when new technologies are considered, as discussed in the following section.

Summary

For each of the different basic technologies different realizations exist. This allows some flexibility for system developers in combining different solutions. For example, a DRM system can be implemented as an open source system. Nevertheless, this flexibility provides difficulties, especially for the compatibility of the different DRM systems.

From a security point of view, it has to be considered that the whole system is as secure as its weakest part. This does not only address the individual components but also their interactions. Thus, components, as well as their integration, have to be chosen carefully.

DRM TECHNOLOGIES

If everything seems under control, you're just not going fast enough.

Mario Andretti

Table 1. Principle characteristics of watermarking and fingerprinting schemes are summarised.

	Watermarking	**Fingerprinting**
Development	Has to be developed for individual media types	
Availability	Audio, video, images, 3D, music scores	Various media types with a focus on audio, video, images
Alteration	Content is altered by embedding	Not necessary but registration in database necessary instead
Registration	Not necessary (cf. alteration)	Prior to identification
Attacks	Vulnerable	Limited vulnerability (perceptual features)
Capacity	Varying on content (minimum requirement should be 64bit for creating a link)	Indirectly in the content and the method's ability to discriminate content
Infrastructure	Depending on application (can be implemented as an independent solution)	Needed (connection to a database)

In this section, the principle components of DRM systems are described. This general principle is more or less underlying each implementation of a DRM system. From an operational point of view, the typical parties involved are the content owner who distributes content, the customer or the consumer who purchases content, and a clearinghouse that manages licenses. For simplicity, we assume that the content owner is also the content distributor, which is not generally the case. If this is not the case, the relationship between the content distributor and the content owner might influence the DRM architecture, as content can be exchanged between these two parties at different security levels.

General Aspects

Sellers of traditional goods benefit from online shops as they are accessible without any time constraint. Product information as well as purchase related information can be made available. But not only traditional goods can be sold on the Internet. Especially content providers of digital content have a general interest in, and strongly benefit from, online distribution. Several advantages can be identified including:

- **Availability:** 24 hours a day and 7 days a week
- **Reduced costs:** Not only complete collections are sold
- **Try before buy:** Customers can have a prelistening/preview
- **Customer relationship:** Direct contact to customers like personalized offers, increased feedback possibilities, and so forth.
- **Reduction of shipping costs:** No physical goods have to be distributed
- **Reduction of storage costs:** Only a digital storage solution is need

Content providers deserve the protection of their content. Therefore content is encrypted before its distribution. As a result, this encrypted content can be distributed in various ways, including download or e-mail transmission. For content usage and rendering, a license is needed, which can be stored locally or on a remote server. This license is used by the client device or client player for decrypting content.

Today's protection solutions, however, are device dependent. Therefore content usage and rendering is device dependent. This is a general problem today. Customers do not want to be restricted by protection solutions. As most available distribution platforms strongly restrict customers (e.g., content can be rendered only on one certain device) in the numerous customers' view DRM is the acronym for digital restriction management.

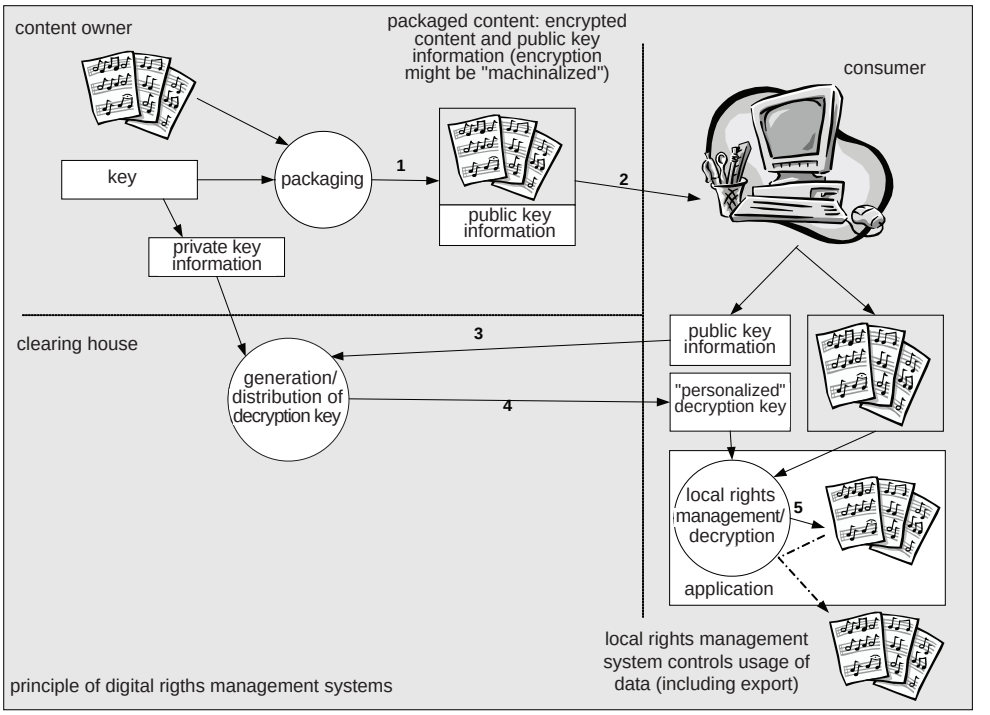
When analyzing a system's security, different assumptions have to be made as described by Arnold et al. (2003). These assumptions include the knowledge of the attacker. It is quite difficult to estimate this, as software patches ("hackz" or "crackz") can be often downloaded from the Internet ("break once, run everywhere"—BORE). These patches allow even users with almost no knowledge to circumvent certain protection mechanisms. Furthermore, the applied security solutions can be secure while the runtime system is not secure at all. This allows potential hackers to successfully attack the runtime system while not interfering with the applied security solutions. Thus, a secure system is vital for the security of content. Therefore the hardware industry is targeting at secure devices[20].

DRM Architecture

The general architecture of a DRM system is shown in Figure 7. Primary technology components can be identified (according to):

- Packagers[21] collect license information, metadata, and the requested content in encrypted files. These secure files are called packages, containers, and so forth.

Figure 7. This figure illustrated the principle architecture of a DRM system. The important aspect is that content is always encrypted outside of the DRM environment. Whenever an application wants to access information, the local rights management system is called to decrypt the content, and it also influences the functionality of the application. Thus a user cannot access any data outside the DRM system.



- Controllers[22] realize local rights management systems at the client side and are responsible for authentication of the device or the user, decryption of the content, provision of access to the content, and sometimes for financial transactions.
- License servers or clearinghouses are run by third trusted parties, and create and distribute licenses that grant access to the requested content via the local rights management system and can also contain access conditions.

As described by Rosenblatt et al. (2002), a broader definition of DRM encompasses everything that can be done to define, manage, and track rights to digital content. Thus further elements are included in this definition:

- Business rights (or contract rights) are typically associated with some content in certain scenarios. For example, the right to use a certain audio sequence in a commercial spot might be granted while processing of the audio sequence is prohibited.
- Access tracking, or tracking of operations on content, provides valuable information for content providers, even if they do not charge for access to the content. This information also helps to improve business models or relationships to customers.
- Rights licensing is an important issue, especially when content can be modified and is redistributed. Yet, technical solutions are strongly limited, for example, when the modification is translation.

Content Owner

Whenever content is distributed, this content is encrypted. The encrypted content is transmitted to the customer with public information about the encryption key. This process is called "packaging" and is step "1" in the figure 7. One has to keep in mind that this is just a simplified view on the package and the packaging itself. Typically, rights information is also included in the package to permit or to restrict certain types of operations, certain operations intervals, or the amount of operations, as well as additional product information (metadata). This rights information is stored in the license.

Content is stored in a repository. Either the repository is built within the DRM solution or it is part of a content management system (CMS). If a CMS stores the content and the packaging system is not able in managing arbitrary file types, the storage format of the content must be chosen according to the capabilities of the packaging respective the DRM system.

Again, it is also necessary to protect the content distribution system. Typical attacks might come from the Internet. These attacks can be faced by a well-configured firewall. Yet, attacks from users with physical access to the content distribution system are also possible. Thus, a certain level of trust in the people working at the content distributor's side is also necessary.

After packaging, content is distributed to the consumers. This can be done using different transmission channels. The Internet via download is probably the most common channel, but also, a transmission via e-mail, floppy disk, or CD is possible. The transmission processes is indicated in step "2."

Consumer

Whenever the consumer receives the content, it is initially useless as it is encrypted. Thus, the corresponding decryption key is needed. Of course it is possible to distribute the keys directly to the customer. Yet, this would reduce the security of the system drastically. Therefore, a local rights management component is responsible for this task. This component requests the keys from the clearinghouse (step "3"), which is a third trusted party. To increase the security, this local rights management component is unique and can be identified. Thus, personalized keys are sent to the local rights management component (step "4"), which makes them only useable for one certain local rights management component.

After receiving the decryption key the content is decrypted. For security reasons, neither the decryption key nor the decrypted content[23] must be stored locally.[24] Therefore, a strong connection between the local rights management component and the application rendering the content is necessary, as content exchange is not only possible via files but also through other channels, like the clipboard or via screenshots.

We would like to stress that the previously described functionality is not restricted to a personal computer. It can also be deployed in other devices. But certain ,for example, mobile consumer devices will result in certain requirements on the complexity of the involved algorithms as their computational power is weaker, and the usability of devices is directly correlated to the execution speed of certain operations.

One important aspect, not only when dealing with mobile devices, is the problem if the DRM solution should also be functional in an offline environment. This requirement increases security threats considerably.

Clearinghouse

The clearinghouse enables the consumer to render the content. The minimalist version transmits "personified" decryption keys to consumers. A more sophisticated version considers licensing issues: The valid content usage period or the amount of rendering. The clearinghouse is also

able to initiate financial transaction, for example, when pay-per-use is demanded in the license.

Rendering Applications

As described, a strong connection between the local rights management is necessary. Rosenblatt et al. (2002) distinguish different rendering applications: stand-alone, plug-in, and java rendering applications:

- Stand-alone rendering applications allow a maximum control of the content. Yet, this advantage has to be paid with several drawbacks: First, the software has to be distributed to the consumers. Second, the consumer has to install the proprietary software on his hardware. Generally, users prefer ready-to-use solutions. They do not want to be bothered with technical details.
- Plug-in rendering applications are common solutions that integrate themselves into existing software. As a direct consequence, the functionality of the "hosting" software is augmented. In the case of DRM plug-ins, it is able to render an increased number of files types. Of course, the plug-in has to modify the "hosting" software's behavior to control data exchange and to avoid any content leakage.

Unfortunately these solutions have to be developed for each hardware platform. From a content distributor's point of view, *Java* combines the advantages of stand-alone and plug-in applications, and additionally throws away the hardware dependency, as Java programs are not run directly on the microprocessor but are executed on a simulated processor, which is called the Java Virtual Machine (JVM). DRM solutions implemented in Java can be run on every processor for which a JVM exists.[25] Today this is the case for most Web browsers. Although Rosenblatt et al. (2002) raise the problem of incompatibilities, an efficient platform-independent DRM solution is currently addressed, for example, by Sun Microsystems.

Security Issues

The main purpose of a DRM solution is the protection of content respective of its license conform usage: content security. Security is always related to certain assumptions. For example, the described assumption of the technical skills of an attacker. But other security issues are directly related to the user and the involved hardware and software platforms.

Digital rights management systems for general content distribution scenarios require the identification of the user. For example, this is very important for secret information exchanged within a company. Similarly, an identification of a customer is important in the music distribution scenario as consumers can be seen as a business partner. For the business transaction, a credit card number might be sufficient. Practically, content usage cannot be limited to the person who purchased it. Thus, information about the person rendering the content is necessary. This information may be a simple e-mail address, a user ID, a password, or other personal information. In other application scenarios, biometric identification systems are used. For example, one can think of personalized mobile devices with biometric sensors: a "lost" mobile device is useless for its "finder." Yet, simple biometric solutions—and these are all current solutions which can be integrated into mass products for monetary reasons—can be easily fooled. Other solutions exist, like the "typewriting style," and are considered by music distribution solution providers. For identification, other possibilities include digital certificates (created by a third trusted party) or smart cards.

Besides user identification, device identification plays an important role. This can be done, for example, by a unique identification number or by the media access control (MAC) address.[26] The advantage of using the MAC address instead

of the IP address is the fact that IP addresses can be dynamic addresses, and also IP addresses can correspond to multiple users.

Device identification is not sufficient at all. One aspect that is generally neglected is the device integrity. The device integrity includes hardware as well as software integrity, which is very difficult. First, hardware and software are under total control of a user.[27] But even if the customer is trustworthy, "external influences" like Trojan horses might violate the device's integrity.

The problem of the device integrity is addressed by the "trusted computing" activities like the Trusted Computing Group (TCG, https://www.trustedcomputinggroup.org/), Trusted Computing Platform Alliance (TCPA, http://www.trustedcomputing.org/), or "Next Generation Secure Computing Base" (NGSCB, http://www.microsoft.com/resources/ngscb).

A trusted environment is typically assumed as a precondition. Yet, this is difficult to achieve, especially whenever the hardware and software cannot be fully controlled, which is usually the case whenever a consumer owns a device.

The trusted computing idea, which is supported by the most hardware and software players, aims to a standard for a PC with increased security level. Although this goal is very important in commercial scenarios (e.g., document security, information assurance…), such a standard is ambivalent for consumers. The danger is that control of individual hardware is transferred from the hardware owner to other parties, like the software vendor implementing the operating system or the content industry in general.

While this is interesting for content distributors, consumers might neglect this standard as from their point of view, the system is less trustworthy, and what is even more important, somebody has to pay for the additional components. The resulting trusted devices will not allow access to decrypted data, for example, through debugging software, will not start modified software, and they will also control the input and output devices.

Integrating DRM Systems with Existing Systems

Although DRM systems can be used as standalone solutions, it is more fruitful to combine DRM systems with other systems to maximize their common benefit. As DRM systems manage content access they can be used whenever content is involved. Thus DRM systems address the complete content "life cycle" and related tools or systems, including:

- Content Creation Tools
- Content Management Systems (CMS)
- Customer Relationship Management (CRM)
- Web publishing systems
- Access Control
- And so forth

In companies, a certain workflow process is established. As modifications of an existing workflow process is very expensive or maybe not possible, deploying DRM systems must not result in any change. This is even more important when techniques or solutions are applied for the protection of content. The protection level of some protection technology is time dependent; it might depend on the time and effort attackers spent in breaking it. If some content requires the highest protection level, the involved protection technology must be updated regularly. Therefore, changes of the workflow process are not manageable. But DRM systems must not only fit in the workflow process, but should also support it.

The general interest in DRM system is reflected in Microsoft's ambitious goal to include a complete DRM system within their operating system. However, there are strong concerns about the integration of DRM functionality in operating systems as discussed, for example, by Anderson (2003, 2004) and the Electronic Frontier Foundation (EFF, http://www.eff.org/).

Content Creation and Management

In business application scenarios dealing with content creation and management, DRM technology can be integrated in the content creation tools and the content management system. The main motivation for this is that content is always stored together with metadata. This metadata may include contract rights or licensing rights.

As an alternative, rights metadata can be created by a manual input. Yet, manual input is expensive as well as error-prone. Thus, a DRM system allows the automation of metadata creation and guarantees its consistency, even if compound works are created. For example, an audiovisual presentation might contain several individual images, video sequences, songs, and speech, which have their individual rights. Content creation and authoring software involving a DRM system can automatically deal with these rights issues, and also solve problems when extracts of such a kind of audiovisual sequence are created.

Besides the storage of rights in a DRM system, fingerprinting and watermarking technologies can link media to the corresponding set of rights. Thus, even a link between the rights and the rights is possible when a media break happened. These content management systems (CMS) integration issues are addressed by Rosenblatt and Dykstra (2003a, 2003b).

Web Publishing and Customer Relationship Management

Deployment of DRM solutions in consumer-related areas typically involve the sales of digital content. This has to be done via an online catalogue or portal. DRM solutions provide the necessary technologies to achieve different business models that better suit the wishes of customers. These business models may include subscription-based services, free time-limited trials, or pay-per-rendering, and can be chosen independently for different customers.

Further improvements are possible when DRM technologies are integrated with customer relationship management systems. Therefore, the offers can be chosen, exactly matching the customers' behavior. For example, whenever a customer purchases a rendering right for a certain content, free time-limited trial rights can be created for related content. Also, the prices for products can be adapted to the usage, allowing a subtle change between pay-per-rendering and subscription based services, for example, as addressed by Rosenblatt and Dykstra (2003a, 2003b).

Access Control

Access control is a desired criterion for content providers and contet owners. Yet, this is not a desired criterion for customers, as they generally do not accept any restrictions on content they purchased. Additionally, access control might interfere with privacy, as discussed by Cohen (2002) and considered by the EU in "Digital Rights—Background, Systems, Assessment."

This is different from companies that want to keep confidential material within their domain. Thus, enterprise content management (ECM) can be regarded as an application that very strongly demands efficient rights management systems. NGSCB and TCG lay the necessary foundation for a secure environment within business applications.[28] As the computers involved in this area are under the control of one administrator, the security assumptions within this scenario are different from the previous scenario. Also DRM systems do not interfere with privacy in this scenario. But DRM systems might interfere with other laws, as access to information can be limited to a certain time interval. And not to forget that also companies, like users, want to define when content is accessible. Both will not accept restriction imposed by third parties, which includes hardware and operating system developers.

Examples for Conflicts between Security and Consumer Issues

Here, two examples are given how the security intention of DRM systems can interfere with consumer issues and sometimes even with law. These incompatibilities can be created artificially for the protection of content, for example, in the case of Macrovision. Also, a proprietary format can result in incompatibilities, for example, in the case of Sony.

- Macrovision's video copy protection system (http://www.macrovision.com/) is a popular product for protecting video content. It was originally developed for video home systems (VHS). An additional copy protection signal is inserted in the part of the video signal used to control the TV (vertical blanking interval and extra synchronization pulses). The idea is that this kind of noise does not interfere with the screen representation of the visual content. But if a video cassette recorder (VCR) tries to make an exact copy of the video signal containing the copy protection signal, it will fail drastically (for the visual content). Interestingly, this idea worked fine with TVs produced before Macrovision's protection system was developed. However, with the growing numbers of DVD devices, people found out that for some TV sets, the only possibility is to connect the DVD to the screen via the VCR. Although the signal is passed only through the VCR in this case, some activate their video-scrambling chip, leading to the same distortion as descried earlier. Additionally some TVs are not capable of resolving the Macrovision signal due to the bas synchronisation pulses. And the Macrovision's video copy protection system delayed the development of DVD players with a progressive[29] output signal.
- Not only Apple with iTunes insists on proprietary format and unrevealed applicaton programming interfaces (APIs); also does (or did) Sony. BETAMAX is an example where Sony failed with a proprietary format. Sony founder Akio Morita saw the reason in the missing license possibilities, resulting in the fact that the inferior VHS system reached a critical mass. Besides BETAMAX, MicroMV was another proprietary format that was abandoned, although it was introduced recently in 2002. Another proprieraty format is ATRAC3, Sony Minidisc format. Even the memory sticks used in Sony's digital cameras and PDAs are proprietary.

Incompatibility, as it regularly has happened with Macrovision's video copy protection system, is also an issue for the protection of music CDs. Incompatibilities interfering with a trouble-free enjoyment led to the creation of "negative lists" like UnCD (cf. http://www.heise.de/ct/cd-register/). Proprietary formats are always critical. For example, Microsoft Office documents can hardly be exported to software products produced by other vendors. But for protected content, this is even more critical. On the one hand content encryption parameters have to be known. On the other hand the representation is unknown.

Potential problems can be foreseen not only in the case of a proprietary format: Customers will not be happy if a proprietary format is abandoned. But furthermore, there is one aspect that is never considered by industry. Archives must provide access to content, even after years of its creation and distribution. They will fail poorly. Today it is already a problem to access content in old formats stored on outdated devices, as hardware and software for access is missing. This problem will become much more severe with encrypted content whose decryption keys are not accessible anymore.

Further Criteria

In addition to the previously mentioned characteristics that should be considered, further criteria that should be evaluated include:

- **Degree of protection:** As described before, one has to be clear who the potential attackers are and what kind of attacks can be performed. On the one hand, typical customers should not be able to break the system. On the other hand, commercially oriented product pirates should not be able to attack a DRM-system successfully. Both groups can be distinguished by their knowledge and the available tools. Especially for commercially oriented pirates there is almost always a possibility to break the security of a DRM-system, as they expect a monetary benefit. Firstly, they can hire professional security experts[30] who have the necessary knowledge. Secondly, they can buy the necessary equipment. Therefore, a huge variety of attacks can be performed, including attacks penetrating hardware.[31]
- **Known attacks:** Considering the security of DRM-systems, one important question is: Are there any known attacks against the complete DRM-system or against individual components? As any DRM system is as strong as its weakest component, one has to consider this aspect carefully. In some cases the weakest points might even be humans. For example, if people are interested in hacking a server, sometimes the easiest way might be to bribe its administrator for creating security holes, which can be exploited easily. Besides the knowledge of successful attacks, details about the attack itself are important, including the effort and the consequences. On the one hand a successful attack might take a lot of time, involve expert knowledge, and has to be performed on each digital item. On the other hand an exploit of the system might have been discovered, resulting in a software patch available on the Internet. Balancing the negative effects, it is obvious that the second kind of attack is much more severe and might almost destroy the security of the whole DRM-system. This is not a theoretic or potential threat. The software industry experienced that hackers[32] react fast and sometimes publish new exploits within a few days in the Internet.
- **Usability:** Any hindrance to content access caused by DRM systems directly reduces user acceptance. Thus, an ideal DRM-system should not be experienced by any user unless a user has the intention to violate the licensing conditions. Unfortunately, this is practically not realizable, as operations cannot be qualified in (current) DRM-systems. For example, a DRM-system cannot distinguish between a legal[33] format conversion and an illegal one. As a consequence, a DRM system prevents all operations that potentially lead to a violation of licensing conditions.[34] One of these effects is the denial of copying in DRM-systems, leading to strong impacts on accustomed ease of handling content.
- **Compatibility:** An important aspect of usability is compatibility. The rendering of content should not be limited to a specific class of devices. This is a major obstacle of today's DRM systems. Although there are ongoing standardization issues, no compatible DRM exists yet.
- **Current popularity:** Obviously, content owners and distributors are in favor of DRM as it supports them in protecting content. Due to restrictions imposed by DRM systems and due to privacy issues most users are against DRM systems. However, the success of iTunes shows that DRM solutions, which are not as restrictive or obvious, are acceptable for customers.

- **Future trends:** DRM is a relatively young issue and one which addresses not only rationality, but also feelings. Also little experience was gained so far with its application. The development of new DRM systems with different kinds of restrictions will show what will be acceptable for customers, practically.

Summary

Obviously, DRM covers a wide range of technologies and also touches legal issues, which makes this topic quite complex. DRM's main purpose is the protection of content. Content security is a very important issue, although not easy to achieve. As a consequence, operating systems implement more and more functionality needed for the protection of content. Their initial implementations, TCG and TCPA, experienced strong resistance, for example, as privacy and traditional content usage are not guaranteed. Nevertheless, there are a lot of hardware systems, especially laptops and mobile phones, that already include the basic functionality. This is almost unknown to most users as this functionality is not used yet.

As explained before, user acceptance is very crucial for the success of business models as well as for the success of DRM systems. Among the important aspects is compatibility and traditional content usage. This does not mean that customers will not accept DRM. Apple shows with its iTunes Music Store—it has a less restrictive content protection policy—that customers indeed are actually willing to pay for DRM-protected content. Nevertheless, iTunes is under strong discussions due to its restrictive DRM solution, from a legal point of view, for example, as in Norway or in France.

TECHNOLOGICAL POSSIBILITIES AND PRACTICAL LIMITATIONS

In dreams and in love there are no impossibilities.

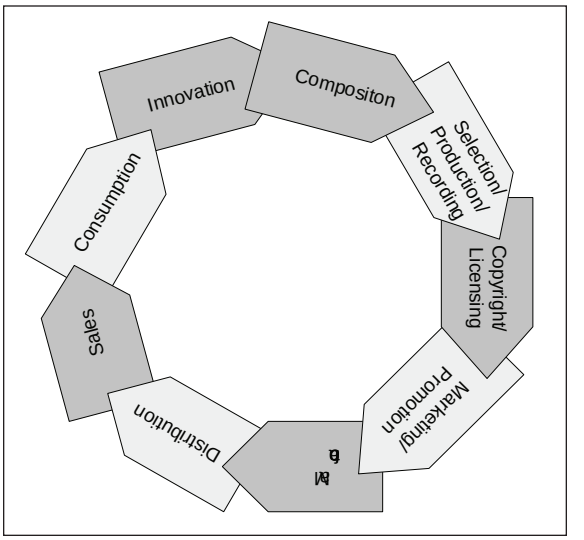
Janos Arany

Content distribution depends on the underlying business model. Therefore, the objective of this section is to describe how DRM can support different business models. For further information, the reader is suggested to consult Rosenblatt (2002), Dubosson-Torbay et al. (Dubosson-Torbay, Pigneur, & Usunier, 2004) or Bockstedt et al. (Bockstedt, Kauffman, & Riggins, 2005).

Traditional Music Industry Value Chain

A simplified traditional music industry value chain is shown in Figure 8. Its purpose is to exemplify the influence of digital representation and DRM

Figure 8. The traditional music value chain involves different players for composition, selection, production, and recording, copyright and licensing, marketing and promotion, manufacturing, distribution, sales, consumption, and innovation.



on content distribution. For specific content like audio or sheet music, the content value chain varies. Different chain links can be identified in the general content value chain:

1. Composition
2. Selection, production, and recording
3. Copyright and licensing
4. Marketing and promotion
5. Manufacturing
6. Distribution
7. Sales
8. Consumption
9. Innovation

This value chain can be considered as a value circle: Existing content is typically the starting point for innovation, which is already the case for the traditional value chain.[35]

Digital Music Industry Value Chain

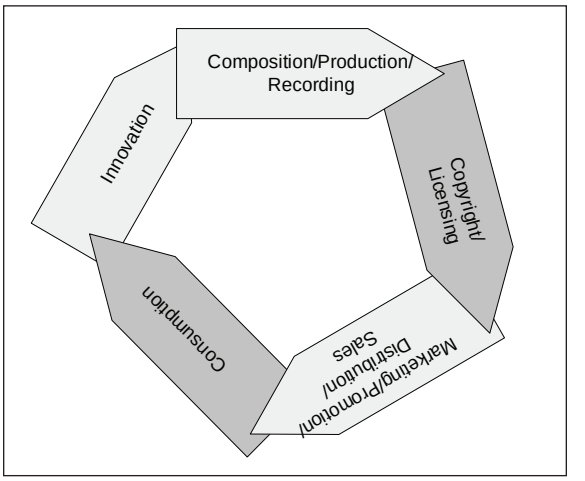
The digital content representation strongly influences the production and distribution of music. Music has become or is becoming a digital good. As a consequence of the digital representation of music, the processing and distribution possibilities are extended. These new possibilities not only blur the boundaries in the traditional content value chain. The result is a downsized content value chain, as shown in Figure 9:

1. Composition, production, and recording
2. Copyright and licensing
3. Marketing and promotion, distribution, sales
4. Consumption
5. Innovation

Influence of DRM on the Digital Music Industry Value Chain

In Figure 9 we showed the influence of digitalization. DRM provides further potentials

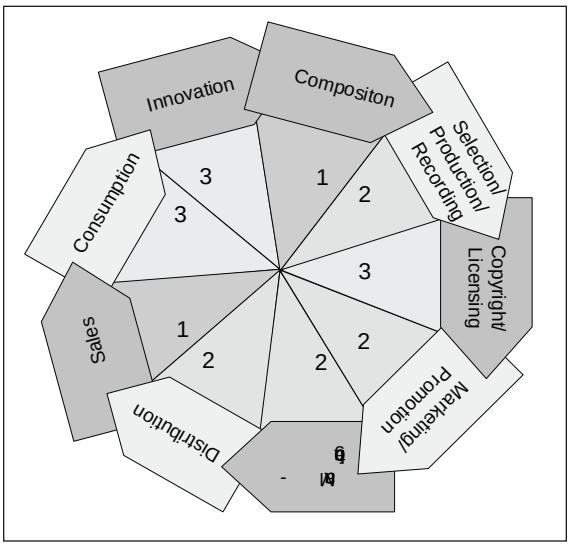
Figure 9. The digital music industry value chain is downsized through the possibilities of digital content representation and its distribution. From content and rights owners' point of view, this allows more efficient content distribution to customers. Furthermore, due to the loss of hardware costs and the capabilities of the computers, more tasks can be performed with less equipment. Obviously, the cycles are rigorously shorter than in the traditional content value chain.



for content creators and content owners: DRM enables control on content usage. Thus content consumption and innovation can be influenced or controlled with DRM. As the basis for DRM are digital (representation of) license, copyright and licensing can be simplified, also as shown in Figure 10. This is so far ongoing development. For example in the AXMEDIS (http://www.axmedis.org) project, the simplification of licensing issues is a central aspect.

Traditionally, a content creator benefits from the composition of the new content and its sale or derived contents thereof. Due to the digitalization, the content creator benefits as selection, production, and recording, marketing and promotion, manufacturing, and distribution is easier and more strongly connected. DRM furthermore has potentials in copyright and licensing, consumption, and innovation. This is shown in Figure 10.

Figure 10. Influence of digitalization and DRM on the content value chain: Traditionally, a content creator benefits from the composition of the new content and its sale or derived contents thereof (labeled with "1"). Due to the digitalization the content creator benefits as selection, production, and recording, marketing and promotion, manufacturing, and distribution is easier and more strongly connected (labeled with "2"). Due to DRM, potentials can be exploited in the areas copyright and licensing, consumption, and innovation (labeled with "3").



Business and License Models

Different business and corresponding license models can be implemented considering the potentials of digital distribution and DRM. The common ones are listed. Further business and license models are described by Dubosson-Torbay et al. (2004). They focus on the influence of emerged P2P legal distribution of free content and illegal distribution of copyright protected content:

- License per content and license per collection of content (paid downloads).
- License per rendering (pay-per-view, pay-per-listen, …).
- License for a specific number of renderings or for a specific time frame.
- Distributed retail by distributing content to retailers (super distribution).
- Super distribution in P2P environments.
- Subscription based services.
- License according to usage.

Licensing Rights

Instead of distributing the content, the rights can be "sold." Of course digital rights management also should address these business models, as selling rights involves rights transactions among commercial organizations.

Summary

DRM provides the possibility to implement a different business model, for example, based on the content consumption. The success of DRM-enabled business models is decided by the customers' acceptance. Furthermore, a successful DRM solution supports business models and does not require the business models to be adapted to the DRM system.

Practically, it does not matter if the business models are adopted models from the physical world, new business models, or a combination of both. Naturally, the business models must be derived from the needs of the customers, which is a well-known fact from the physical world. Yet, new technologies provide possibilities for further products and services.

It should be also stressed here that only few content creators' primary income in the music domain is based on the commercial distribution of their music or songs. A huge part lives on secondary sources, like live performances or ring tones. Thus, limiting access to content is not in the interest for these artists. As a consequence, each artist has to decide carefully the personal

pros and cons before applying existing DRM technologies to her/his content.

TOWARDS THE CUSTOMERS

All government—indeed, every human benefit and enjoyment, every virtue and every prudent act—is founded on compromise and barter.

Edmund Burke

Although music labels were strongly in favor of CD copy protection technologies, more and more music labels do not apply CD copy protection technologies anymore. Users did not appreciate this technology that they practically realized as an obstacle.

This is something content owners and distributors have to understand. Maybe the following statement explains the users' restrictions best: "I will accept your DRM protected content if you accept my DRM-protected money!" Thus, new solutions are needed considering the customers' requirements and objections. Similarly, the decision of a French court against DVD copy protection shows that new approaches have to be developed. In this section, we summarize recently developed and on-going developments that better consider customers' requirements.

Obviously, DRM and content protection is still a hot topic. But as shown by Apple's iTunes MusicStore, and some others that want to get a share from Apple's success, users potentially accept content protection technologies. Nevertheless, there are ongoing developments that focus on the user requirements that have not been sufficiently considered yet.

Interoperable DRM

One of the major reasons customers are very reluctant against it is compatibility, respectively, the lack of compatibility. "Buy once, listen anywhere and anytime" is not possible with current DRM systems. Actually, this was one of the reasons why customers reject copy protected CDs: They cannot be played on all consumer devices (including computers).

In the digital world the situation is even more complex: Having different storage carriers, different media encodings and recently, different DRM containers dramatically increases the difficulties in exchanging content.[36] And for ordinary users this is a big issue: they are interested in enjoying content; they neither want to be bothered with any technical detail nor with any problem.

As described before, there are ongoing efforts that address compatible DRM, which we include:

- **CORAL:** The Coral Consortium is "a cross-industry group to promote interoperability between digital rights management (DRM) technologies used in the consumer media market. The Consortium's goal is to create a common technology framework for content, device, and service providers, regardless of the DRM technologies they use. This open technology framework will enable a simple and consistent digital entertainment experience for consumers." (http://www.coral-interop.org/)
- **DMP:** The Digital Media Project is a not-for-profit organization with the mission to "promote continuing successful development, deployment and use of Digital Media that respect the rights of creators and rights holders to exploit their works, the wish of end users to fully enjoy the benefits of Digital Media and the interests of various value-chain players to provide products and services." (http://www.dmpf.org/)
- **ISMA:** The Internet Streaming Media Alliance's goal is "to accelerate the adoption and deployment of open standards for streaming rich media content such as video, audio, and associated data, over Internet protocols." (http://www.isma.tv/)

- **OMA:** Open Mobile Alliance is similar to ISMA, which is not limited to a specific technology, but is in favor of a certain range of applications, OMA wants to promote the usage of mobile devices. Its mission is "to facilitate global user adoption of mobile data services by specifying market driven mobile service enablers that ensure service interoperability across devices, geographies, service providers, operators, and networks, while allowing businesses to compete through innovation and differentiation." (http://www.openmobilealliance.org/)

Unfortunately, there is a competition between the different standardization activities. It is very interesting to see how the different standardization activities developed during the last years a good summary is given by Rosenblatt (2004, 2005, 2006, 2006b).

In the mobile area, OMA so far succeeded in the sense that new mobile phones already integrate OMA-DRM. The new extension of OMA addressing entertainment devices at home thus seems promising, although OMA seems to have slowed down. DMP seems to lack a user group and electronics companies, and still is in the specification phase in contrast to CORAL, which already showed how to build interoperable DRM, for example, with Microsoft Windows Media DRM. CORAL would benefit from Apple that should publish their interfaces. In addition to these activities, Sun initiated a project called Open Media Commons (http://www.openmediacommons.org/). For this open source "DRM/everywhere available" (DReaM), no license cost should limit its usage. If a reliable security level can be guaranteed in open source, DRM has to be proven in the future. The first implementation, however, is already limited to IPTV. The first implementations are expected in 2007.

Less Interfering DRM

Developing interoperable DRM systems is only one direction for improving the usability of DRM-protected content. Another direction is not to develop new formats and devices, but to use existing formats and embedding customer information. This has been described before as transaction watermarks.

One example for this is the Light Weight Digital Rights Management (http://www.lwdrm.com/), which cooperates with consumers. It distinguishes between locally bound and signed content in which user certificates are enclosed. As these user certificates link the content to users, users are expected to be very reluctant in distributing this personalized content.

The philosophy behind this is quite simple, like the car's license plate: In principle, the users can do anything with their content. There is no restriction unless the content is not distributed to others. If such personified content occurs on illegal file sharing networks, the responsible user can be identified and blamed for this. This is similar if a driver violates a speed limit and can be punished for this.

Besides the usage aspect—traditional content usage is not prevented—user awareness is increased: Users become responsible for the content they purchase.

New Developments in P2P-Networks

There are also ongoing developments that try to commercialize the distribution in P2P networks. One approach is to distribute DRM-protected content via P2P networks. Examples are Napster (http://www.napster.com/) for audio content or in2movies (http://www.in2movies.de/in2movies/) for video content. Besides these traditional approaches, new developments that avoid classical DRM protection methods have been developed.

Music2Share

Kalker et al. (Kalker, Epema, Hartel, Lagendijk, & van Steen, 2004) presented a solution where each client integrates fingerprinting and watermarking techniques. By this the system allows the identification of the content exchanged. Simplified, before a client is allowed to receive content, his/her request is evaluated from a trusted third party to identify, if he/she already licensed the content. If so, the content is transferred to the customer. This allows putting different business models on top of existing P2P networks. After receiving the content, the content is encrypted or put in a DRM container, together with the licensing information retrieved from the TTP. Again, users experience restrictions when content is DRM protected.

The "Social Distribution Network" (SDN)

User awareness and user responsibility are the key concepts of the so-called social distribution networks (SDNs), as presented by Schmucker (2004) and Schmucker and Ebinger (2005a).[37]

The authors motivate them by the following train of thoughts:

- Perfect content protection is not possible. There are always security leaks like the analog hole. This has to be considered when efforts are spent for the development, integration, or usage of DRM. This means that at least specialists "unprotect" DRM protected content.
- Content exchange via the Internet cannot be controlled.
- Unknown artists do not care about DRM and illegal distribution, as their primary income is not from record sales but from performances, ring tones, and so forth. Nevertheless, they deserve a platform for the exchange, and especially the promotion of their content. This platform requires the possibility to stop the delivering of content upon request.
- Users are not bad. Costumers will not redistribute illegal copies to harm artists. Instead they are willing to pay a reasonable amount of money for music that is unrestricted in its usage.
- Costumers will support artists by promoting content. Additionally, they will create additional metadata. This additional metadata further promotes artists' content.

Commercial DRM-protected content is less attractive and competes with illegally distributed content that does not have any usage limitations.

A distribution system that integrates consumers is most attractive if it is based on P2P networks: "Let them make work (pay) for you" is the direct benefit for artists. P2P users pay for storage and bandwidth. A P2P-based architecture was presented by Schmucker and Ebinger (2005b) that considered those issues mentioned earlier,and provides a framework for the secure exchange of unprotected content.

Everybody who is registered can distribute content in the P2P network. User awareness is a central issue. It is achieved by the feedback to the customers: If the content that should be distributed is already registered as nonsharable, the user gets a message and cannot distribute this content. In contrast, if this content is not registered, the user has to register it to him/her. After this registration, content can be distributed freely within the P2P network.

Due to the fact that content itself is not DRM protected, this distribution network is very interesting for the promotional distribution of content, especially for relatively unknown artists.

New Licensing and Business Models

In addition to the technological developments, new licensing and business models are being

developed. In general, they are not limited to a certain distribution channel. Some of them are presented here.

Super Distribution

The Potato-System (http://www.potatosystem.com/) and WeedShare (http://www.weedshare.com/) are typical super distribution business models. The users who distribute content of artists get money for the distribution if the recipient buys the music.

Creative Commons

Creative Commons provides a set of copyright licenses free for public use. (http://creativecommons.org/). One can select a suitable license easily, based upon some preferences/selections. This license model is related to GNU Public License (GPL) and EFF Open Audio License. Some projects evolved from creative commons like science commons (http://sciencecommons.org/) or iCommons, which aims at a "united global commons front" (http://icommons.org/).

Fisher's License Model

Fisher's license model is based on a low-rate subscription. Fisher (2004) showed in detail how such a model could work, and Orlowski (2004) analyzed it.

German Academic Publisher

German Academic Publishers (http://www.gap-portal.de/) is developing a new model for the design and administration of electronic publications. It is especially interested in the scientist's need for "Open Access."

Summary

Customers' acceptance is the most critical issue of content protection technology. Thus, current developments try to solve today's systems' flaws. An interesting aspectsis to raise consumers awareness and their responsibility, while incorporating them in the distribution of content either buy online reputation systems or by using their resources. There is a huge potential in the user communities, as examples in other areas show. Interestingly, the successful content distribution platforms, like YouTube (http://www.youtube.com) or BeSonic (http://www.besonic.com), incorporate users in producing, rating, and ranking content, but not in its distribution.

Return on investment (ROI) or return on capital employed (ROCE) plays an important role for nonprofit as well as for commercial organizations. Both judge solutions' and technologies' benefit according to their ROI. ROI is even more important if venture capital is involved, or a company is listed in the stock market.

One problem of general security solutions is to measure this return on investment. Costs for new technology can be measured accurately, but what about the saved amount of money? The problem return on security invest (ROSI) is currently highly debated.

As DRM belongs to the category of security mechanisms, the same problems arise here. Yet, additional return on investment factors can be identified:

- Cost saving resulting from electronic stock.
- Cost saving resulting from traditional content delivery (including packaging or redelivery).
- Flexibility and scalability of the online solution (business model can be changed or adopted easily).
- Improved services to customers (e.g., availability, transmission speed, …).

- Improved customer relationship management.
- Improved monitoring capabilities.
- Improved brand loyalty.

Again, these returns are difficult to measure. But it should be clear that the benefit of a modern combined distribution and DRM solution is far beyond the return on investment by protection of content after the point of purchase.

Bomsel and Geffroy (2004) characterized the economic characteristics of DRM. Two main economic functions of DRM are defined: content protection and versioning. Then, the mechanisms of DRM adaptations are analyzed. Network effects are considered in this analysis. Interestingly, this analysis distinguishes between open and dedicated networks for content delivery. For example, mobiles are considered a one-way dedicated channel for content delivery. Although the analysis identified the DRM's significance for content delivery, three obstacles are identified for DRM's roll-out in open architectures (in contrast to dedicated networks).

1. "The current economic model of Internet access … does not favour non free open services."
2. "In broadband networks, pay content distribution conflicts with the network effects pulling the roll-out."
3. "Content owners may push alternative networks."

According to Bomsel and Geffroy (2004), one consequence is that "broadband open networks may give priority to two way communication services … rather than to pay models for digital media content distribution."

THE FUTURE SOLUTION

Again we would like to emphasize that DRM systems integration within other solutions, like content management systems, will increase their benefit most. Aspects that cannot be neglected in the design and decision process include the business models and the workflow process. Depending on the customers, a less-restrictive DRM, for example, lightweight DRM, might even be the better solution.

The change in the hardware and software solutions for content distribution is still ongoing. Their tendency is going into the direction of "trusted computers" or "trusted devices." Independently how long it takes to achieve these ambitious goals, which have some advantages and also some drawbacks, customers have to accept these solutions and also to pay for them. This development towards "trusted devices" is highly debated, as it seems to endanger the right for "free speech" and free information exchange. Future solutions of trusted computers might allow the customers a more flexible key management away from hardware keys stored in machines, but with keys stored in smart cards. This will also be influenced by other standardization activities like DMP, OMA, CRF, or CORAL.

So what the future might bring is hard to say, as history already told us that not the best solution must succeed in the long run.[38] Nevertheless, GartnerG2 and the Berkman Center for Internet & Society presented "five possible scenarios for copyright law applicable to digital media in the United States" in the report "Five Scenarios for Digital Media in a Post-Napster World." These scenarios predicted losses and gains for consumer values, and costs and revenues for content owners, artists, technology CE vendors, and Internet service providers:

1. The no-change scenario is based on the assumption that DMCA is still enforced: "This scenario is the least likely to play out, as the entertainment industries are not likely to sit still and see their business models slowly destroyed. Media companies have already attempted to address piracy via legal, regulatory and technology solutions. They will continue to pursue solutions to what they perceive as an attack on their traditional business models. However, it is likely that the no-change scenario will prevail in the immediate future as efforts so far have yielded minimal results and piracy is still widespread."
2. The taking property rights seriously scenario is based on the assumption that content owners and providers strongly succeed in protection of their IPRs. As a result, the gains for content owners and artists will increase together with increased cost on the overhead with violation prosecution. Technology and Internet service providers will gain marginally and the consumers will be the losers. "This scenario certainly plays to the interests of those in the media industry and copyright holders who would seek to maintain existing business models based on complete control of the content. However, it is probably the one scenario that best illustrates the chasm separating content owners/media companies from large segments of the consumer population. It is also the scenario that, if realized, would most emphatically underscore the regional differences in intellectual property laws and enforcement. "
3. The effective technology defense scenario assumptions are that content will be distributed physically and digitally. Content is copy protected while still meeting consumers' needs. It also includes the assumption that copy protection is an ongoing cycle, which is indeed the case. "This scenario can be described as 'technology rescues the content industries from wanton copyright piracy.' However, the technological challenges are compounded by the numbers of increasingly tech-savvy consumers around the world. There is very little margin for error and the transition to universal copy protection must be relatively quick. Otherwise, media companies and artists may find that large numbers of consumers are seeking digital content from sources other than traditional music labels, movie studios and publishers." The difference between this scenario and the second scenario is that the second scenario assumes legal reforms while the third scenario assumes technological changes.[39]
4. The compulsory license scenario assumes that the current copyright system is replaced by a system in which the creators and producers of content are compensated by the government in proportion to the "consumption" frequency. "While this scenario has its own risks—giving a government entity significant discretionary power and assuring the virtual annihilation of the physical retail market—the potential for reducing litigation, lowering the costs of enforcement and eliminating the incentive for an ongoing encryption 'arms race' make it very attractive."
5. The utility model scenario considers digital content as a public utility. Regulations are enforced by a federal regulatory body. Concerning the estimated effects, this scenario is most interesting. "Of all five scenarios presented here, this one countenances major legal, business and consumer behaviors changes. From a technology perspective, it is less complicated than might be considered. At least one technology provider currently has an offering that could track content distribution to the end user in much the same way power companies use meter-reading systems. However, music and

> movie producers and their businesses—not to mention conventional retail distribution entities—will be violently opposed. Music and movie producers would see their revenue models altered greatly, with the costs associated with distributing content and usage eliminated."

Thus the "utility model scenario," as described in the report "Five Scenarios for Digital Media in a Post-Napster World" by GartnerG2 and the Berkman Center for Internet & Society, might be the best solution to the current problem of the content owner. But before such a final solution is publicly accepted and established, content providers have to find their individual solutions.

The Evolution of P2P Networks

P2P technology is constantly evolving. Thus, it is interesting to see what will be the next step in this evolution. Biddle et al. (Biddle, *England, Peinado, & Willman,* 2002) from Microsoft expect that "interconnected small world darknets" will come into existence. In this network type, small groups of people exchange content within. The different groups are connected through people being a member of several groups. Due to the structure, it is very difficult to control and therefore to stop exchange within it.

This development has gone far beyond from shared file systems. Friendster (http://www.friendster.com/) indicated this development and was also one of Napster's success factors according to Roettgers (2004): Communities were part of Napster. And solutions like AllPeers (http://www.allpeers.com/) enable content sharing with friends and family directly via a Web browser plug-in.

Current P2P rather satisfies the users' need for anonymity due to the legal pressure. "The Free Network Project" (http://freenet.sourceforge.net/) is probably the P2P network addressing this requirement for anonymity best. It "is entirely decentralized and publishers and consumers of information are anonymous. Without anonymity there can never be true freedom of speech, and without decentralization the network will be vulnerable to attack." Although its intended aim is to stop censorship, it can also be misused for illegal content exchange. And other developments are ongoing as well.

This is in contradiction to the trust-based communities like Friendster. Incorporating trusted communities in P2P networks will result in "social P2P networks." They will exist in parallel to the anonymous P2P networks.

Besides legal measures and court trials, the evolution of P2P technology also resulted in the evolution of technology, which can be used as countermeasures against illegal distribution. One popular approach is the introduction of wrong or manipulated files in the P2P networks. Fingerprinting solutions are used in commercial and educational environments as well as in the P2P-networks, for example, in Napster. Also, the Internet traffic is statistically analyzed to identify potential P2P applications. Early 2007, YouTube announced that it will also deploy fingerprinting technologies for IPR protection.

Instead of fighting P2P networks, advantages are more and more realized. For example, P2P networks can be used for promotion of new material. Additionally, P2P networks are a distribution channel, as recently used by George Michael. Furthermore, content is exchanged between people. Statistical analysis can be performed on this exchange with the aim to identify new trends in music, as done by Big Champagne (http://www.bigchampagne.com/).

The Analogue Hole

DRM developments aim at trusted systems: Systems that cannot be "manipulated" by users. But there is a general flaw in trying to protect content: the analogue hole. There must be a version of the content that can be accessed by humans: Images and videos are visualized and audio is played.

The resulting signal, which is intended for human spectators or listeners, can also be recorded in a digital format again. In this case, watermarking, if it is sufficiently robust, presents the only possibility to trace back the origin of the leakage. This might not be possible somewhere in the future, when electrodes directly stimulate the brain. But in the meantime, an analog signal has to be presented to the ears or the eyes.

There are already companies trying to address exactly this problem. For example, DarkNoiseTechnologies[40] tries to insert a "dark noise" in the signal, with the aim that recorded versions of the signal suffer severe distortions.

Examples for Music Distribution

Within this chapter, an overview about the technological aspects of audiovisual protection was provided. As already discussed, protection solutions have to form a unity with other technologies, for example, CMS or CRMS, in distribution systems. Different online music distribution services are described in several reports at the Interactive Musicnetwork (http://www.interactivemusicnetwork.org).

All of them apply DRM technology. Although in the case of Apple iTunes MusicStore, the DRM is less restrictive and allows burning CDs. Maybe this is one of the reasons why Apple's approach is the most successful so far.

Summary

In spite of all efforts, the current situation is controversial: From a consumer's perspective, any usage restriction reduces content value. From a content owner's perspective, digital content cannot be distributed without any protection. Hence, a compromise must be found between the consumers and the content owners. So far DRM is not generally accepted.

Nevertheless, DRM has the potential to be part of this compromise. For example, the versioning, as identified by Bomsel and Geffroy (2004), provides interesting opportunities. However, the content owners have to be aware that the new possibilities in usage control that are offered by DRM, are neither wanted nor enjoyed by customers today. It seems that in early 2007, the music industry realizes the problems caused by current DRM systems. But alternatives that are attractive for content owners and customers so far are still missing.

Thus, each content owner has to perform an objective analysis on the individual DRM usage. This analysis has to consider his/her needs and the needs and the requirements of the customers as well. Only by this, realistic opportunities can be identified in contrast to the hopes and expectations that were initially created by technology developers. Only then, DRM has realistic chances to be a valuable technology.

Not to be forgotten, that technology is often used in different ways, as initially intended by its developers. The potentials and the future usage of DRM might also be different from its initial intensions. Thus, continuously analyzing new application scenarios and technologies that were initially not covered by DRM technology is mndatory.

REFERENCES

Allamanche. E., Herre. J., Hellmuth. O., Froba. B., & Cremer, M. (2001). AudioID: Toward content-based identification of audio material, In *Proceedings 110th AES*, Amsterdam, The Netherlands.

American Library Association. (2005). *Digital rights management and libraries*. Retrieved March 20, 2007, from http://www.ala.org/ala/washoff/WOissues/copyrightb/digitalrights/digitalrightsmanagement.htm

Anderson, R. (2003). *Trusted computing—Frequently asked questions*, August, 2003. Retrieved March 21, 2007, from http://www.cl.cam.ac.uk/~rja14/tcpa-faq.html

Anderson, R. (2004). *Economics and security resource page, 2004.* Retrieved March 21, 2007, from http://www.cl.cam.ac.uk/~rja14/econsec.html

Arnold, M., Schmucker, M., & Wolthusen, S. (2003). *Techniques and applications of digital wtermarking and content protection.* Boston: Artech House.

Arrest in 'Screener' Upload Case. *CBS News*, Los Angeles, January 23, 2004. Retrieved March 21, 2007, from http://www.cbsnews.com/stories/2004/02/18/entertainment/main600881.shtml

Bell, T. W. (2007). *Music copyrights table.* Retrieved March 20, 2007, from http://www.tomwbell.com/teaching/Music(C)s.html

Biddle, P., England, P., Peinado, M., & Willman, B. (2002). The darknet and the future of content distribution. *Microsoft Corporation,* 2002 ACM Workshop on Digital Rights Management. Retrieved from http://crypto.stanford.edu/DRM2002/darknet5.doc

Bockstedt, J. C., Kauffman, R. J., & Riggins, F. J. (2005). The move to artist-led online music distribution: Explaining structural changes in the digital music market. In *Proceedings of the 38th Hawaii International Conference on System Sciences, 2005.* Retrieved from http://misrc.umn.edu/workingpapers/fullpapers/2004/0422_091204.pdf

Bomsel, O., & Geffroy, A. G. (2004). *Economic analysis of digital rights management systems (DRMs).* MediaNet Project Paper, December, 2004. Retrieved from http://www.cerna.ensmp.fr/Documents/OB-AGG-EtudeDRM.pdf

Camp, L. J. (2003). First principles of copyright for DRM Design. *IEEE Internet Computing, 7*(3).

Cano, P., Baltle, E., Kalker, T., & Haitsma, J. (2002). A review of algorithms for audio fingerprinting. *IEEE Workshop on Multimedia Signal Processing.*

Cohnen, J. E. (2002). *DRM and privacy.* Research Paper Series, Research Paper No. 372741, Georgetown University Law Center.

Commission Staff Working Paper, SEC. (2002). *Digital rights—Background, systems, assessment, 197,* February, Brussels.

Copyright Issues for Music. Music Library, University at Buffalo. Retrieved March 20, 2007, from http://ublib.buffalo.edu/libraries/units/music/copyright.html

Cox, I. J., Miller, M. L., & Bloom, J. A. (2002). Digital watermarking. In *The Morgan Kaufmann Series in Multimedia Information and Systems.* San Francisco: Morgan Kaufmann Publishers.

Craver, S. A., Wu, M., Liu B., Stubblefield, A., Swartzlander, B., Wallach, D. S., Dean, D., & Felten, E. W. (2001). Reading between the lines: Lessons from the SDMI challenge. In *Proceedings of the 10th USENIX Security Symposium,* Washington, DC.

Dubosson-Torbay, M., Pigneur, Y., & Usunier, J. C. (2004). Business models for music distribution after the P2P revolution. In *WEDELMUSIC '04: Proceedings of the Web Delivering of Music, Fourth International Conference on Web Delivery Of Music.* Washington, DC: IEEE Computer Society.

EBU metadata specifications. *European Broadcasting Union (EBU).* Retrieved March 21, 2007, from http://www.ebu.ch/en/technical/metadata/specifications/index.php

FBI Arrests Internet Movie Pirate. *FOX News,* January 23, 2004. Retrieved March 21, 2007, from http://www.foxnews.com/story/0,2933,109272,00.html

Felten, E. (2001). *Statement at the fourth international information hiding workshop in Pittsburgh.* Retrieved from http://www.cs.princeton.edu/sip/sdmi/sdmimessage.txt

Fisher, W. (2004). Promises to keep—*Technology, law, and the future of rntertainment.* Retrieved March 23, 2007, from http://www.tfisher.org/PTK.htm

GartnerG2 and the Berkman Center for Internet & Society. (2003). *Five scenarios for digital media in a post-Napster world.* Retrieved March 21, 2007, from http://cyber.law.harvard.edu/home/research_publication_series

Greek, D. (2004). *RIAA crackdown shows signs of success.* Retrieved March 20, 2007, from http://www.vnunet.com/vnunet/news/2124044/riaa-crackdown-shows-signs-success

Grimm, R. (2005). *Privacy for digital rights management products and their business cases.* Virtual Goods Workshop at IEEE Axmedis 2005, Firence, Italy, December.

Haitsma, J., van der Veen, M., Kalker, K., & Brueker, F. (2000). *Audio watermarking for monitoring and copy protection.* ACM Multimedia Workshops.

Iannella, R. (2001). Digital rights management architectures. *D-Lib Magazine, 7*(6). Retrieved March 20, 2007, from http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/june01/iannella/06iannella.html

Intellectual property. *Wikipedia.* Retrieved March 20, 2007, from http://en.wikipedia.org/wiki/Intellectual_property

International Standard Text Code. *ISO/TC 46/SC 9 WG3 Project 21047.* Retrieved March 20, 2007, from http://www.nlc-bnc.ca/iso/tc46sc9/wg3.htm

Kalker, T., Epema, D. H. J., Hartel, P. H., Lagendijk, R. L., & van Steen, M. (2004). Music-2Share—Copyright-compliant music sharing in P2P systems. *Proceedings of the IEEE, 92*(6), 961-970.

Katzenbeisser, S., & Petitcolas, F. A. P. (Eds.). (2000). *Information hiding: Techniques for steganography and digital watermarking.* Boston: Artech House.

Kerkhoffs, A. (1883). La Cryptographie Militaire. *Journal des Sciences Militaires,* 9th series, 5-38,161-191.

Lewis, G. J. (2003). *The German e-music industry.* Leonardo de Vinci, Faculty of Business Administration, University of Applied Sciences, Dresden, Germany. Retrieved from http://imec.hud.ac.uk/imec/iMEC%20GermanFINALreport110902.pdf

Library and Archives Canada. *International Standard Music Number.* Retrieved March 20, 2007, from http://www.collectionscanada.ca/ismn/

Madden, M., & Rainie, L. (2005). *Music and video downloading moves beyond P2P.* Pew Internet & American Life Project, Retrieved from http://www.pewinternet.org/pdfs/PIP_Filesharing_March05.pdf

Man nabbed for uploading Oscar "screener." *CNet News.com,* February 22, 2007. Retrieved March 21, 2007, from http://news.com.com/2061-10802_3-6161586.html

Marion, A., & Hacking, E. H. (1998). Educational publishing and the World Wide Web. *Journal of Interactive Media in Education, 98*(2). Retrieved from http://www-jime.open.ac.uk/98/2

Menezes, A., Oorschot, P. & van Vanstone, S. (1996). *Handbook of applied cryptography.* CRC Press.

Network Working Group, The Internet Society. *RFC 1737: Functional Requirements for Uniform Resource Names.* Retrieved March 20, 2007, from http://www.ietf.org/rfc/rfc1737.txt

Network Working Group, The Internet Society. (1995). *RFC 1736: Functional Recommendations*

for Internet Resource Locators. Retrieved March 20, 2007, from http://www.ietf.org/rfc/rfc1736.txt

Network Working Group, The Internet Society. (1998). *RFC 2396: Uniform Resource Identifiers (URI): Generic Syntax*, Retrieved March 20, 2007, from http://www.ietf.org/rfc/rfc2396.txt

Orlowski, A. (2004). Free legal downloads for 6$ a month. DRM free. The artists get paid. We explain how… *The Register*, February, 2004. Retrieved March 23, 2007, from http://www.theregister.co.uk/content/6/35260.html

Orlowski, A. (2005). French court bans DVD DRM. *The Register*, 26/04/2005. Retrieved March 20, 2007, from http://www.theregister.co.uk/2005/04/26/french_drm_non/

Peer-To-Peer. *Wikipedia*. Retrieved March 20, 2007, from http://en.wikipedia.org/wiki/Peer-to-peer

Peticolas, F. (2006). *Stirmark benchmark 4.0*, 06 February 2006. Retrieved March 22, 2007, from http://www.petitcolas.net/fabien/watermarking/stirmark/

Rhodes, R. (1899). *Pianola copyright ruling cited*. Retrieved March 20, 2007, from http://mmd.foxtail.com/Archives/Digests/200406/2004.06.05.03.html

Roettgers, J. (2004). *Social networks: The Future of P2P file sharing*. Retrieved from http://freebitflows.t0.or.at/f/about/roettgers

Rosenblatt, B., Trippe, B., & Mooney, S. (2002). *Digital rights management—Business and technology*. New York: M&T Books,

Rosenblatt, B., & Dykstra, G. (2003a). *Technology integration opportunities*, November, 14, 2003. Retrieved March 21, 2007 from http://www.drmwatch.com/resources/whitepapers/article.php/11655_3112011_3

Rosenblatt, B., & Dykstra, G. (2003b). *Integrating content management with digital rights management—Imperatives and opportunities for digital content*. Lifecycles. GiantSteps, Media Technology Strategies , Technical Report. Retrieved from http://www.xrml.org/reference/CM-DRMwhitepaper.pdf

Rosenblatt, B. (2004). *Review: DRM Standards*, January 5, 2004. Retrieved March 21, 2007 from http://www.drmwatch.com/standards/article.php/3295291

Rosenblatt, B. (2005). *Review: DRM Standards*, January 6, 2005. Retrieved March 21, 2007 from http://www.drmwatch.com/standards/article.php/3455231

Rosenblatt, B. (2006). *Review: DRM Standards*, January 2, 2006. Retrieved March 21, 2007 from http://www.drmwatch.com/standards/article.php/3574511

Rosenblatt, B. (2006b). *Review: DRM Standards*, December 27, 2006. Retrieved March 21, 2007 from http://www.drmwatch.com/standards/article.php/3651126

Rudish, J. (2005). Illegal file sharing lawsuits dismissed; Emory withholds defendants' names. *The Emory Wheel Online*, April 15, 2005. Retrieved March 20, 2007, from http://media.www.emorywheel.com/media/storage/paper919/news/2005/04/15/News/Illegal.Filesharing.Lawsuits.Dismissed.Emory.Witholds.Defendants.Names-1647598.shtml

Schmucker, M. (2004). *Enlightening the DarkNet*. IST-Workshop, November, 15th, 2004, Den Hague, Netherlands.

Schmucker, M., & Rantasa, P. (2005a). *Ein soziales Verteilungsnetzwerk—Beispiel eines alternativen Distributionsmodells*. CAST-Workshop, February 2nd, 2005, Darmstadt, Germany.

Schmucker, M., & Ebinger, P. (2005b). *Promotional and commercial content distribution based on a legal and trusted P2P framework*. CEC 2005. Seventh IEEE International Conference on E-Commerce Technology.

Scmucker, M. (2005c). *Protection of coded music*. The Interactive Musicnework. Retrieved from http://www.interactivemusicnetwork.org/documenti/view_document.php?file_id=1195

Schneier, B. (1996). *Applied cryptography* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

SMEF Data Model. *British Broadcasting Corporation*. Retrieved March 21, 2007, from http://www.bbc.co.uk/guidelines/smef/

Templeton, B, (2007). *10 big myths about copyright explained*. Retrieved March 20, 2007, from http://www.templetons.com/brad/copymyths.html

Walton, T. (2003). *Golden age of free music vs. Copying is stealing.* Retrieved March 20, 2007, from http://www.theregister.co.uk/2003/08/06/golden_age_of_free_music/

Wang, X., Feng, D., Lai, x., & Yu, H. (2004). Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD. *Cryptology ePrint Archive, Report 2004/199*. Retrieved from http://eprint.iacr.org/2004/199

Wang, X., Yin, X. L., & Yu, H. (2005). *Collision search attacks on SHA1*, February 13, 2005. Retrieved from http://theory.csail.mit.edu/~yiqun/shanote.pdf

XML and Digital Rights Management (DRM). *Cover Pages* (hosted by OASIS). Retrieved March 21, 2007, from http://xml.coverpages.org/drm.html

ENDNOTES

1 For example, in the U.S., the copyright for a new composition lasts for the lifetime of the composer and additional 70 years after his death.

2 Generally, user anonymity is not given. All users can be traced back by their IP and with the support of the ISP. It is very important for wireless-LAN (WLAN) operator to secure access to their WLAN against misuse, as these operators are responsible for any kind of misuse.

3 Fair use is a statutory exemption to the copyright law.

4 Interestingly, the existing solutions analysed by Camp, which included copy protection as well as circumvention technologies, only partially fulfilled these requirements.

5 The analogue mass production ensured that a document survives unaltered and can be located.

6 The publishers' and broadcasting investment results in a careful selection of content.

7 An information collection is available at RIAA (http://www.riaa.com/).

8 Freenet can be summarized as a decentralized network of file-sharing nodes tied together with strong encryption and further technology, which allows anonymous users.

9 Besides confidentiality, other relevant aspects might be authentication, integrity, or copyright protection, which has to be addressed using different techniques.

10 Yet this distinction is somewhat hazy as block ciphers can be used as stream ciphers, and vice versa.

11 These algorithms are also known as one-way hash algorithms.

12 Typically a soft-hash or perceptual hash value is embedded, which is another term for fingerprinting.

13 A watermark in an image or audio can be used to start a plug-in in a Web browser for automatic linking the content to a certain Web site.

14 More or less a perceptible watermark in music scores already exists: the copyright information.

15 Asymmetric watermarking schemes have been developed as well, but they have some drawbacks.

16 Blind watermarking schemes are also called public watermarking schemes.

17 These are also called private watermarking schemes.

18 For some media types (e.g., text or music scores), embedding a watermark directly in the content is very difficult. Here, the representation is modified, which effects robustness against conversion and potentially the perceived quality.

19 Sometimes even the term "passive watermarking" is used, which we consider as misleading as no mark is embedded.

20 These secure devices are also called trusted devices, reflecting the assumption that trusts in the security of the systems can be provided.

21 Typically, content servers provide this functionality.

22 Controllers are sometimes also called "DRM controllers."

23 Local storage of the decrypted content depends on the business model. Some business models might allow this. Some might only allow local storage with poor quality (e.g., strongly compressed audio files).

24 Current solutions implement a key cache to increase the systems flexibility. But this also increases the systems vulnerability.

25 Of course the processing power must be sufficient.

26 The MAC address is a unique value associated with a network adapter, and are also known as hardware addresses or physical addresses.

27 At least this is the case nowadays. This might change in the future if the "trusted computing" initiatives succeed. Yet consumers have to pay for this technology and they do not only benefit from it.

28 Other possible consequences, for example, software monopoles, have to be considered carefully, and a thorough observation is necessary to avoid negative effects to economy.

29 Typically, a TV set uses interlacing, which combines two "half"-images into one. DVD players are capable of producing a progressive output signal; one complete frame at once.

30 Due to their motivation "semiprofessional" hackers can be considered as a big threat.

31 Thus, people interested in the protection of their content have to be aware that the higher the monetary benefit is, the bigger is the potential danger that an attacker has commercial interests.

32 Sometimes this kind of hacker is also called crackers.

33 For example, archiving or providing access to visually impaired people are most often allowed by law.

34 Interestingly in the physical world, such a practise is not manageable or allowed. For example, knives are still sold although people can potentially be hurt.

35 There are different opinions on this on other areas, for example, as expressed by Marion and Hacking (1998). For the digital value chain this might be more evident, as content available in digital format can be processed more easily. Thus the cycles are rigorously shorter.

36 This is a general problem especially for digital archives, even without DRM: Will any device to read the content stored on a specific carrier be available in 20 or 30 years?

[37] This, and the resulting architecture, was a direct result of the discussion within the MUSICNETWORK when looking for a possible decentralised solution for unknown musicians.

[38] One example is the success of VHS against Betamax or Video 2000, although VHS was inferior against its competitors.

[39] Although the second and the third scenarios are linked due to the results, the difference is in how rights are established and enforced.

[40] DarkNoiseTechnologies was bought by SunnComm (http://www.sunncomm.com).

This work was previously published in Interactive Multimedia Music Technologies, edited by K. Ng and P. Nesi, pp. 283-324, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.10
Modular Implementation of an Ontology-Driven Multimedia Content Delivery Application for Mobile Networks

Robert Zehetmayer
University of Vienna, Austria

Wolfgang Klas
University of Vienna, Austria

Ross King
Research Studio Digital Memory Engineering, Austria

ABSTRACT

Today, mobile multimedia applications provide customers with only limited means to define what information they wish to receive. However, customers would prefer to receive content that reflects specific personal interests. In this chapter we present a prototype multimedia application that demonstrates personalised content delivery using the multimedia messaging service (MMS) protocol. The development of the application was based on the multimedia middleware framework METIS, which can be easily tailored to specific application needs. The principle application logic was constructed through three indepdent modules, or "plug-ins" that make use of METIS and its underlying event system: the harverster module, which automatically collects multimedia content from configured RSS feeds, the news module, which builds custom content based on user preferences, and the MMS module, which is reponsible for broadcasting the resulting multimedia messages. Our experience with the implementation demonstrated the rapid and modular development made possible by such a flexible middleware framework.

INTRODUCTION

Multimedia messaging service (MMS) has not achieved a similar market acceptance and customer adoption rate as *short message service* (SMS), but is nevertheless one of the primary drivers of new income streams for telecommunication companies and is, in the long run, on the way to becoming a true mass market (Rao & Minakakis, 2003). It provides new opportunities for customised content services and represents a significant advance for innovative mobile applications (Malladi & Agrawal, 2002).

Until now, however, mobile operators have failed to deliver meaningful focused mobile services to their users and customers. Telecommunication companies have made considerable investments (license, implementation costs) into third generation (3G) mobile networks but have not yet generated compensating revenue streams (Vlachos & Vrechopoulos, 2004). Customers are often tired of receiving information from which they get no added value, because the information does not reflect their personal interests and circumstances (Sarker & Wells, 2003). The goal is instead to establish a one-to-one relationship with the user and provide costumers with relevant information only. Through personalisation, the number of messages the customer receives will decrease significantly, thus reducing the number of irrelevant and unwanted messages (Ho & Kwok, 2003).

Currently available MMS subscription services (e.g., Vodafone, 2005) allow customers to define what kind of information they want to receive in a very limited way. Broad categories like *Sports, Business,* or *Headline News* can be defined, but there is no generic mechanism for the selection of more specific concepts within a given domain of interest. The personalised and context-aware services demanded by savvy customers require a mediation layer between the users and content that is capable of modelling complex semantic annotations and relationships, as well as traditional and strongly-typed metadata. These will be defining characteristics of next-generation multimedia middleware.

This paper describes the modular development of a mobile news application, based on a custom multimedia middleware framework. The application supports ontology-driven semantic classification of multimedia content gathered using a widespread news markup language. It allows users to subscribe to content within a particular domain of interest and filters information according to the user's preferences. Moreover it delivers the content via MMS. The example domain of interest is the Soccer World Cup 2006 for which a prototypical ontology for personal news feeds has been developed. However, the middleware framework enables mobile multimedia delivery that is completely independent from the underlying domain-specific ontology.

BACKGROUND AND RELATED WORK

Related Work

At this time, there are no readily available systems that combine the power of ontology-based classification, published syndicated content, and a personalised MMS delivery mechanism. There are however a number of proposals and applications that make use of principles and procedures that are similar to those presented in this chapter.

Closely related to the classification aspect of the presented MMS news application are hierarchical textual classification procedures such as D'Alessio, Murray, Schiaffino, and Kreshenbaum (2000). These approaches mostly consider the categorisation of Web pages and e-mails (see also Sakurai & Suyama, 2005) and classify content according to a fixed classification scheme.

Ontologies that can provide classification in the form of concepts and relationships within a particular domain are used by Patel, Supekar, and Lee

(2003) for similar purposes. The idea behind their work is to use a hierarchical ontology structure in order to suggest the topic of a text. Terms that are extracted from a specific textual representation are mapped on to their corresponding concepts in an ontology. The use of ontologies is one step ahead of the use of general classification schemes as they introduce meaningful semantics between classified items. Similar in this respect is the work of Alani et al. (2003), which attempts to automatically extract ontology-based metadata and build an associated knowledge base from Web pages. The reverse method is also possible as demonstrated by Stuckenschmidt and van Harmelen (2001), who built an ontology from textual analysis instead of classifying the text according to an ontology. Schober, Hermes, and Herzog (2004) go one step further by extending the ontological classification scheme from textual information to images and their associated metadata.

Even more closely related to the topics presented in this paper are the techniques employed in the *news engine Web services* application (News, 2005), which is currently under development. It is based on the news syndication format PRISM and ontological classification, and its goal is to develop news intelligence technology for the semantic Web. This application should enable users to access, select, and personalise the reception of multimedia news content using semantic-based classification and associated inference mechanisms (Fernandez-Garcia & Sanchez-Fernandez, 2004).

News Markup Languages and Standards

News syndication is the process of making content available to a range of news subscribers free of charge or by licensing. This section briefly sketches three current technologies and standards in the field of news syndication: RSS, PRISM, and NewsML.

Our MMS application employs RSS feeds in order to harvest news data, due to the volume and free availability of these types of feeds. Of course this would raise serious copyright issues in a commercial application; however, our approach provides an initial proof of concept, allows the harvesting of significant volumes of data for testing classification algorithms, and is easily upgradeable to a commercially appropriate standard, thanks to the modular nature of the system architecture. For this reason, we describe the RSS standard in more detail than the other more commercially significant standards.

Rich Site Summary (RSS)

First introduced by Netscape in 1999, RSS (which can stand for *RDF site summary*, *rich site summary*, or *really simple syndication* depending on the RSS version) is a group of free lightweight XML-based (quasi) standards that allow users to syndicate, describe and distribute Web site and news content, respectively. Using these formats, content providers distribute headlines and up-to-date content in a brief and structured way. Essentially, RSS describes recent additions to a Web site by making periodical updates. At the same time, consumers use RSS readers and news aggregators to manage and monitor their favourite feeds in one centralised program or location (Hammersley, 2003).

RSS comes in three different flavours: relatively outdated RSS 0.9x, RSS 1.0 and RSS 2.0. RSS 2.0 is currently maintained by the *Berkman Center for Internet and Society* at Harvard. On the other hand RSS 1.0 is a *World Wide Web Consortium* (W3C) standard and was developed independently. Thus RSS 2.0 is not an advancement of RSS 1.0, despite what the version numbers might suggest. The line of RSS development was split into two rival branches that are only marginally compatible. The main difference is that RSS 1.0 is based on the W3C *resource description framework* (RDF) standard, whereas the other

types are not (Wustemann, 2004). In our MMS news application scenario the focus is on RSS 2.0 channels, because of their special characteristics relating to multimedia content and the general availability of feeds of this type in contrast to RSS 1.0.

The top level of an RSS 2.0 document is always a single RSS element, which is followed by a single channel element containing the entire feed's content and its associated metadata. Each channel element incorporates a number of elements providing information on the feed as a whole and furthermore item elements that constitute the actual news and their corresponding message bodies. Items consist of a title element (the headline), a description element (the news text), a link (for further reading), some metadata tags and one or more optional enclosure elements. Enclosures are particularly important in the context of multimedia applications, as they provide external links to additional media files associated with a message item. Enclosures can be images, audio or video files, but also executables or additional text files, and they are used for building up the multimedia base of our MMS news application.

Publishing Requirements for Industry Standard Metadata (PRISM)

Publishing Requirements for Industry Standard Metadata (PRISM, 2004) is a project to build standard XML metadata vocabularies for the publishing industry to facilitate syndicating, aggregating and processing of news, book, magazine and journal content of any type. It provides a framework for the preservation and exchange of content and of its associated metadata through the definition of metadata elements that describe the content in detail.

The impetus behind PRISM is the need for publishers to make effective use of metadata to cut costs from production operations and to increase revenue streams as well as availability for their already produced content through new electronic distribution methods. Metadata in this context makes it possible to automate processes such as content searching, determining rights ownership and personalisation.

News Markup Language (NewsML)

News Markup Language (NewsML) is an open XML-based electronic news standard developed and ratified by the *International Press Telec Council* (IPTC) and lead-managed by the world's largest electronic news provider Reuters (IPTC, 2005). According to Reuters (2005), NewsML could revolutionise publishing, because it allows publishers and journalists to deliver their news and stories to a range of different devices including cell phones, PDAs, and desktop computers. At the same time, it allows content providers to attach rich metadata so that customers only receive the most relevant information according to their preferences.

NewsML is extensible and flexible to suit individual user's needs. The goal is to facilitate the exchange of any kind of news, be it text, photos or other media, accurately and quickly, but it may also be used for news storage and publication of news feeds. This is achieved by bundling the content in a way that allows highly automated processing (NewsML, 2003).

Multimedia Messaging Service and Mobile Network Architecture

Multimedia messaging service (MMS) is an extension to the *short message service* (SMS) protocol, using the *wireless application protocol* (WAP) as enabling technology that allows users to send and receive messages containing any mixture of text, graphics, photographic images, speech and music clips or video sequences. High-speed communication and transmission technologies, such as *general packet radio services* (GPRS) and *universal mobile telecommunications system* (UMTS), provide support for powerful and fast

messaging applications (Sony Ericsson Developers Guidelines, 2004).

MMS Network Architecture

An MMS-enabled mobile device communicates with a WAP gateway using WAP transport protocols over GPRS or UMTS networks.

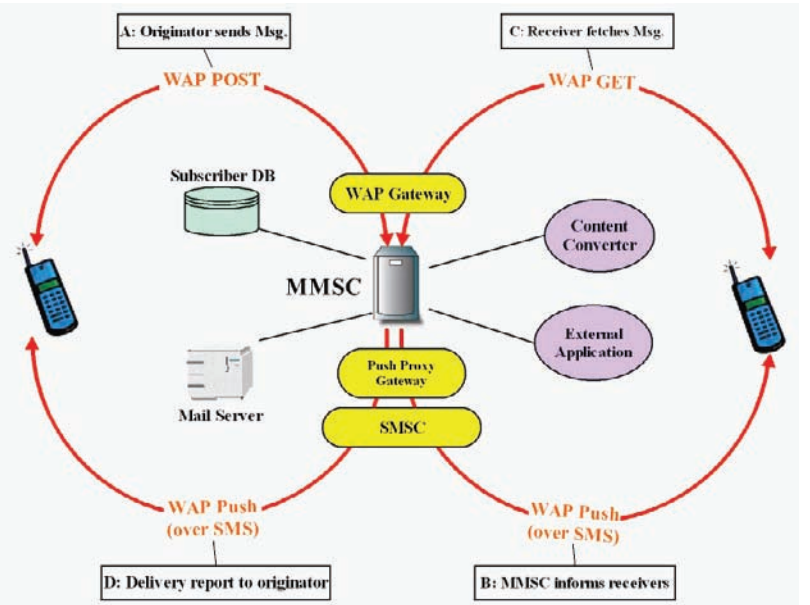
Data is transported between the WAP Gateway and the MMS Centre (MMSC) using the HTTP protocol as indicated in Figure 1. The MMSC is the central and most vital part of the architecture and consists of an MMS Server and an MMS Proxy-Relay. Amongst other functions it stores MMS messages, forwards and routes messages to external networks (external MMSCs), delivers MMS via e-mail (using the SMTP protocol), and performs content adaptation according to the information known about the receiver's mobile phone. This is managed via so-called *user agent profiles* that identify the capabilities of cell phones registered in a provider's network (Sony Ericsson Developers Guidelines, 2004). Leveraging the content-adaption capability of the MMSC is a key feature of our MMS application.

MMS and SMIL

The *Synchronized Multimedia Integration Language* (SMIL) is a simple but powerful XML-based language specified by the W3C that provides mechanisms for the presentation of multimedia objects (Bulterman & Rutledge, 2004). The concept of SMIL as well as MMS presentations in general includes the ordering, layout, sequencing, and timing of multimedia objects as the four important functions of multimedia presentations. Thus a sender of a multimedia message can use SMIL to organise the multimedia content and to define how the multimedia objects are displayed on the receiving device (OMA, 2005).

A subset of SMIL elements must be used (and are used by our application) to determine the presentation format of an MMS message. Listing 1 shows an example SMIL document defining 2 slides (<par> elements), each containing a text,

Figure 1. MMS network architecture (Nokia Technical Report, 2003)



an image, and an audio element, as it would be the case in typical MMS.

MMS Message Structure and WSP Packaging

MMS is implemented on top of WAP 1.2.1 (as of October 2004) and supports messages of up to 100 Kbytes, including header information and payload. In order to transmit an MMS message, all of its parts must be assembled into a multi-part message associated with a corresponding MIME (*multipurpose Internet mail extensions*) type, similar to the manner in which these types are used in other standards such as HTML or SMTP. What is actually sent are so-called MMS *protocol data units* (PDUs). An example of which is shown in Figure 2. In the next step, PDUs are passed into the content section of a *wireless session protocol* (WSP) message, in the case of most mobile networks, or a HTTP message otherwise (Nokia Technical Report, 2003).

One of three possible content type parameters is associated with these content sections, specifying the type of the MMS (Sony Ericsson Developers Guidelines, 2004):

- **Application/vnd.wap.multipart.related:** This type is used if there is a SMIL part present in the MMS. The header must then also include a type parameter application/smil on the first possible position
- **Application/vnd.wap.multipart.mixed:** Used if no SMIL part is included in the MMS
- **Application/vnd.wap.multipart.alternative:** Indicates that the MMS contains alternative content types. The receiving device will choose one supported MIME type from the available ones

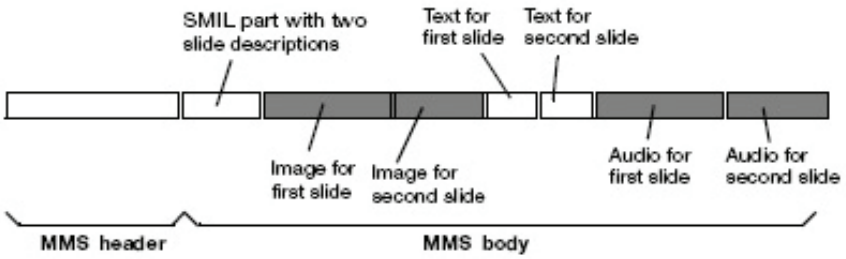
The Multimedia Middleware Framework METIS

The following sections give an overview of the METIS multimedia framework, its generic data model, and methods for the extension of its basic functionality by developing semantic modules and

Listing 1. SMIL XML example

```
<?xml version="1.0" encoding="UTF-8"?>
<smil xmlns="http://www.w3.org/2001/SMIL20/Language">
 <head>
      <meta name="title" content="vacation photos" />
      <meta name="author" content="Danny Wyatt" />
      <layout>
       <root-layout width="160" height="120"/>
       <region id="Image" width="100%"
            height="80" left="0" top="0" />
       <region id="Text" width="100%"
            height="40" left="0" top="80" />
      </layout>
 </head>
 <body>
      <par dur="8s">
       <img src="FirstImage.jpg" region="Image"/>
       <text src="FirstText.txt" region="Text"/>
       <audio src="FirstSound.amr"/>
      </par>
      <par dur="7s">
       <img src="SecondImage.jpg" region="Image"/>
       <text src="SecondText.txt" region="Text"/>
      </par>
 </body>
</smil>
```

Figure 2. Example MMS PDUs



kernel plug-ins. An introduction to the template mechanism that is extensively used in our application is also provided.

System Overview

The METIS framework (King, Popitsch, & Westermann, 2004) provides an infrastructure for the rapid development of multimedia applications. It is essentially a classical middleware application located between highly customisable persistence and visualisation layers. Flexibility was one of the primary design criteria for METIS. As can be seen in Figure 3, this criterion especially applies to the back-end and front-end components of the architecture as well as to the general extensibility through kernel plug-ins and semantic modules. The design as a whole offers a variety of options for the adaptation to specific application needs.

METIS Data Model

The METIS data model provides the basis for complex, typed metadata attributes, hierarchical classification, and content virtualisation. Application developers need only consider their specific data models at the level of ontologies (specified, for example, by RDFS or OWL) which can then be easily mapped to the METIS data model using existing tools. Object relational modelling is handled by the framework and the developer need never concern himself with relational tables or SQL statements.

Figure 3. METIS system architecture

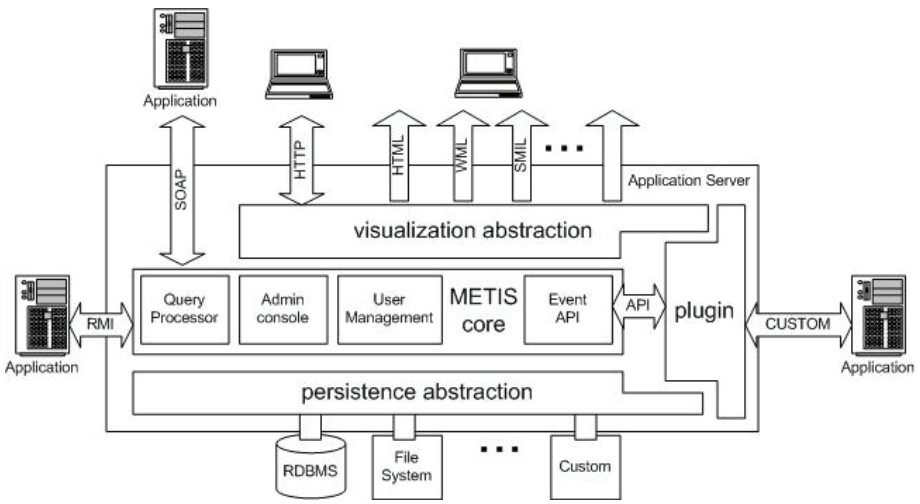


Figure 4. METIS Core Data Model (King et al., 2004)

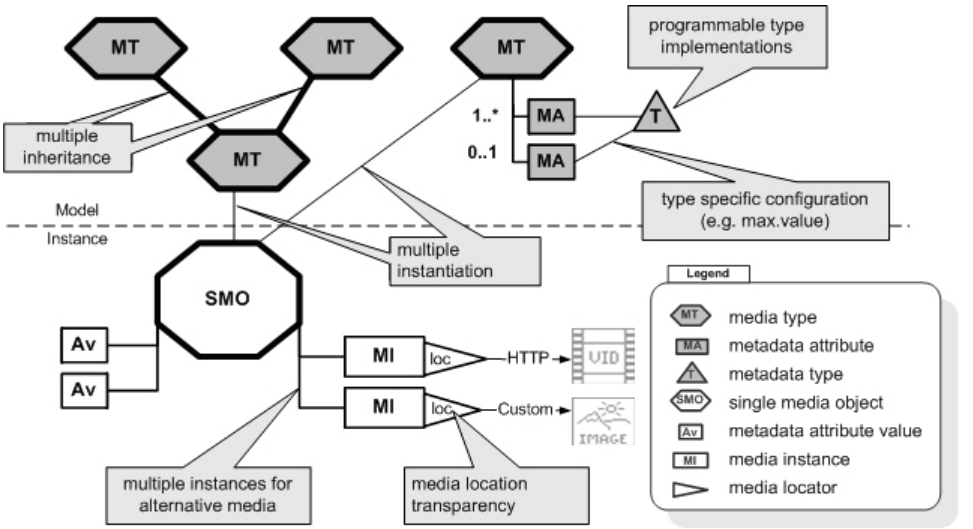


Figure 4 illustrates the basic building blocks of the model and their relationships. Media in METIS are represented as a so-called *single media objects* (SMOs), which are abstract, logical representations of one or more physical media items. Media items are attached to a SMO as *media instances* and connected to the actual media data via *media locaters*, which are in turn a kind of pointer to the data, allowing METIS to address transparently media items in a variety of distributed storage locations such as file systems, databases or Web servers.

As a foundation for semantic classification, media objects can be organised in logical hierarchical categories, known as *media types*. Media types can take part in multiple inheritance as well as multiple instantiation associations. *Metadata attributes* are connected to media types, can be as simple or complex as desired, and can be shared among multiple media types with different cardinalities, default values, and ranges.

Finally, media objects can be connected to each other by binary directed relationships (so-called *associations*). The semantics of these associations are defined by *association types* that are freely configurable within an application domain.

As mentioned previously, there exist simple tools with which domain semantics can be packaged as semantic modules (also called *semantic packs*) that can be dynamically loaded in a given METIS instance and thereby provide the required domain-specific customisations.

Complex Media Objects and Templates

For modelling specific media documents that are made up of several media items, the METIS data model provides *complex media objects* (CMOs). CMOs are quite similar to SMOs when it comes to instantiating media types, taking part in associations and being described by metadata attributes. The crucial difference is that they serve as containers for other media objects, either SMOs or other CMOs. Complex media objects can be rendered in specific visualisation formats by applying the METIS template mechanism (King et al., 2004). A template is an XML representation of a specific multimedia document format (such as SMIL, HTML or SVG), enriched by *placeholders*. When a visualisation of the CMO is requested, these placeholders are dynamically substituted by specific data extracted from the CMO employing

that template, using a format-specific XSLT style sheet. Our MMS application makes use of this template mechanism in order to define the format of MMS messages, by employing the SMIL-based mechanism described in a previous section.

Semantic Modules and Kernel Plugins and the Event Framework

Kernel plug-ins constitute the functional components of an application that extend the basic functionalities provided by the METIS core. These plug-ins not only have access to all customisation frameworks within METIS, but also to the event system, which provides a basic publish/subscribe mechanism. Through the METIS framework, plug-ins can subscribe to certain predefined METIS events and can easily implement their own new application-specific events. This loose coupling between functional extensions provided by the event framework allows large modular applications to be implemented with METIS.

THE MMS NEWS APPLICATION

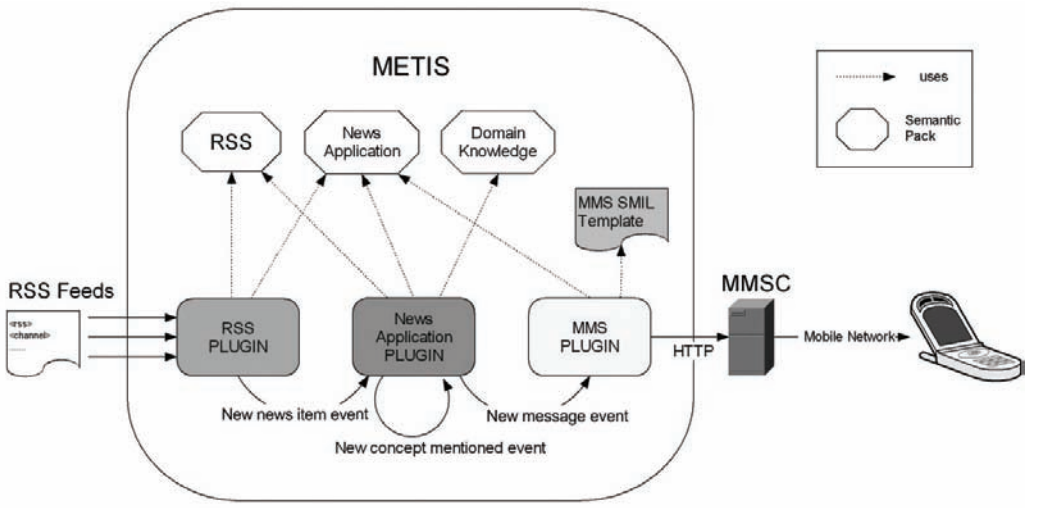
The METIS framework is used extensively to implement our modular application for content delivery in mobile networks. This *MMS news application* illustrates two strengths of METIS: extensibility and fast implementation time.

In order to demonstrate these advantages and the core functionalities, the prototype news application implements a showcase in the wider area of the Soccer World Cup 2006 in Germany. We present an ontology for this domain, which allows a relatively confined set of topics and their relationships to be modelled. However, the system is designed to be as open and extensible as possible and allows mobile multimedia content delivery that is completely independent from the underlying domain-specific ontology.

System Architecture

An overview of principal components of the *MMS news application*'s modular architecture is given in Figure 5. The implementation is split into three functional parts: the RSS import module, the news application module (containing the main applica-

Figure 5. MMS news application architecture (simplified)



tion logic), and the MMS output and transmission module. Each module is implemented as a kernel plug-in, and each module is loosely connected with other plug-ins through the METIS event mechanism.

This approach makes it possible to cleanly separate functionalities into logical modules. It is therefore simple to integrate various functional units into the application's context and substitute existing plug-ins with newly implemented ones whenever changes in the application's environment are required. The interface to which all these plug-ins must adhere is defined by the various events that are issued by components that adopt a given role in the application.

From a high-level perspective, the RSS plug-in takes the role of the multimedia content source that loads multimedia news items into the system. Obviously, RSS would not be the choice for a commercial application; the previously mentioned NewsML and PRISM standards, whose feeds are not normally free of charge, would be more powerful alternatives.

RSS was chosen for the prototype as it allows the demonstration of the essential strengths and advantages of the presented approach with no associated costs. Furthermore it is easy to implement and allows the testing of the whole application on large datasets. In the future, additional multimedia content source plug-ins based on NewsML or PRISM could be quickly developed to replace the RSS plug-in.

The *NewsApplication* plug-in is the core module of the whole application. It integrates the surrounding plug-ins and uses their provided functionalities to create personalised news content. This plug-in itself offers flexibility in the mechanisms used to find topics mentioned in news items as well as in the creation of messages for specific users.

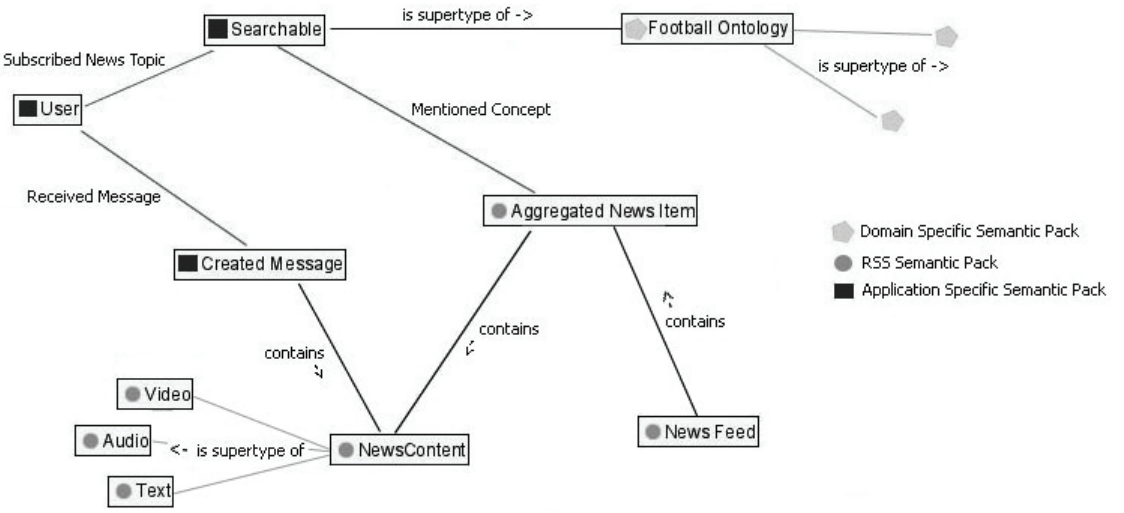
The MMS plug-in fulfils the role of the content delivery mechanism within the *MMS news application* by linking the application to mobile network environments. In the current prototype it is used to send MMS messages via an associated MMSC to a user's mobile handset. Once again, the MMS plug-in offers a variety of extension possibilities and is very flexible when it comes to the system used for the actual MMS transmission. It could be easily substituted by other content delivery plug-ins that target different receiving environments and devices. For example, one might consider a SMS delivery mechanism or a mechanism that delivers aggregated news feeds about certain topics to Web-based news reader applications.

Data Model

The data model and domain-specific semantics of a METIS-based application must be specified through the *semantic pack* mechanism. Semantic packs are to a large extent quite similar to ontologies that define the semantics of specific domains of knowledge by modelling classes, attributes, and relationships. The *MMS news application* is based on three independent semantic packs:

- **RSS semantic pack:** This module maps the RSS 2.0 element and attribute sets to the METIS environment, and supports import from previous RSS versions including RSS 1.0 (without additional modules). Media types included are *news feed*, *aggregated news item*, *news content* with corresponding attributes (e.g., *title*, *description* or *publication date*) as well as general purpose media types such as *image*, *text*, *audio*, and *video* that are all child elements of *news content*. Associations between these elements are defined as well. Generally, this semantic pack is intended to be as independent as possible from the underlying publishing standard that is used, and as extensible as possible in order to facilitate the implementation of other types of import plug-ins
- **News application semantic pack:** This module provides the application-specific

Figure 6. MMS news application semantic pack dependencies



management ontology. It defines media types and metadata attributes that are required by the internal logic of the application in order to store and differentiate between application-specific media objects. Media types in this category are *user, created message,* and *searchable*. A *User* normally subscribes to multiple *searchable* media object instances (SMOs) that are supplied by the domain-specific semantic pack, and associations of type *subscribed news topic* are created between these. Furthermore, associations of type *received message* are instantiated between a *user* and all the *created messages* he has received as a result of his subscriptions

- **Domain-specific semantic pack:** This module constitutes the domain-specific component of the application used for the subscription services and the applied ontology-based classification method. The application's internal logic is completely independent of the domain of interest that is defined by this semantic pack. As a demonstrator, an ontology for soccer was implemented, but additional domains can be implemented and plugged into the existing application with minimal effort

The general dependencies between the three semantic packs and the specific media and association types are presented in Figure 6.

Domain-Specific Semantics and Knowledge Base

The domain-specific semantic pack contains key concepts and their relationships within a specific domain of interest, and defines the structure of a knowledge base containing specific instances of defined classes that must be instantiated. The *MMS news application* is independent of the domain of interest supplied by this semantic pack; any ontology satisfying the basic requirement of having a single parent class from which all other classes are directly or indirectly derived can be loaded into the system and used as a basis for the subscription mechanism.

Domain concepts or classes are stored in the METIS environment as media types. A concept instance is modelled as a SMO of the corresponding concept's media type. In our prototype, all classes are direct or indirect subclasses of the abstract base class *Football Ontology.* Example classes (media types) are *Field Player, Trainer,*

National Team, *Club*, and *Referee*. Furthermore, an application-specific media type *searchable* is included, which provides the required *search term* metadata attribute. This search term enables the textual identification of the instance through the presently simple algorithm based on matching regular-expressions. Domain associations form the basis for the semantic classification algorithm as they relate concepts (i.e. classes) and establish meaningful relationships between them.

Instances (e.g., *David Beckham*) of concepts (e.g., *Field Player*) can be included within the semantic pack itself or defined via the *news application's* user interface. The only constraint that instances within an imported ontology must satisfy is that they must supply at least one identifying search term string attribute for the ontology-based classification mechanism.

Every instance added to the system becomes visible to end-users, who can then subscribe to specific concept instances and receive MMS messages associated with them.

In the case of our prototype, a knowledge base of about 250 instances and their associations was developed in approximately 4 hours. This suggests that it is possible to implement other domains of interest and to adapt the whole application to other application scenarios in a reasonably short time.

Module Integration and Event Mechanism

The RSS plug-in provides all RSS-related mechanisms. RSS news feeds typically contain news items that contain the actual messages. Whenever an item is added and stored, the RSS plug-in informs all interested system components of this fact via a *new news item event*. The only subscriber to this event in the current architecture is the news application plug-in, which is subsequently activated. It searches the new news item for occurrences of domain-specific concept instances (e.g., the instance *David Beckham*) contained in the domain-specific knowledge base. Whenever such an occurrence is found, a *new concept mentioned event* is issued. The news application then attempts to find subscribers to the discovered concept instance (i.e., users who want to receive messages about it) as well as subscribers of associated instances. Associated concept instances in this respect mean instances that are directly connected to the discovered concept through a relationship in the domain-specific ontology. If a user has chosen to receive messages from related instances (by default, a user would receive messages only directly related to the subscribed concept), he will also be added to the set of found users. As an example, consider a subscriber of the instance *English national team* who also chooses to receive messages from related concept instances; he would, for example, also receive messages about *David Beckham*, because *Beckham* is a member of that team. In this case, the user would be an *indirect subscriber* to the Beckham concept instance.

Whenever direct or indirect subscribers are found, the plug-in creates a new CMO (of type *created message*) containing various SMOs such as a news text, suitable images, video or audio items. It is important to note that this newly created message is not a one-to-one translation of the news item contained in the RSS feed. The news application searches the multimedia document base and tries to find media instances that are associated with the discovered concept instance and may be suitable for the newly created message.

The architecture is designed to be as open and extensible as possible. Implementations of new algorithms for ontology-based classification and the associated message-creation mechanism can be easily upgraded within in the application.

Having assembled this message, a *new message event* is issued and the MMS plug-in, as a subscriber to this event, sends the message as a MMS to the users' mobile phones. Outgoing messages are formatted using the METIS template mechanism in conjunction with a predefined MMS

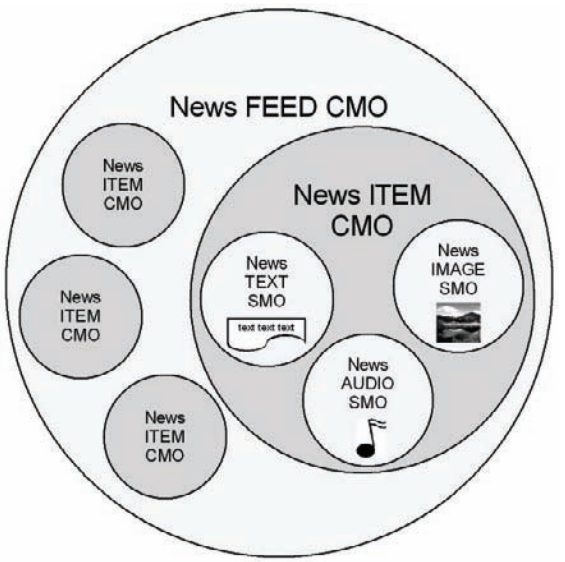
SMIL template. The application could be easily extended to allow users to choose from a variety of templates and define the final format of their received messages.

RSS Import

The RSS import plug-in fulfils the role of an RSS input parser and news aggregator that manages multiple RSS feeds simultaneously and makes their content available to the other components of the application.

Using media types and attributes specified in the RSS semantic pack, the RSS plug-in maps feeds to corresponding METIS media objects by parsing these and extracting media and metadata. In general, a feed is represented as a METIS CMO as depicted in Figure 7. The *FEED CMO* (type: *news feed*) can incorporate several *News ITEM CMOs* (type *aggregated news item*), which in turn include multiple *media SMOs* (subtypes of *news content*) that map RSS media enclosures included in the feed.

Figure 7. News FEED complex media object containing CMOs and SMOs



By regularly searching and updating the stored feeds, a multimedia document base is gradually constructed over time.

The RSS plug-in also functions as a common RSS newsreader and aggregator by providing an HTML visualisation of the created *News FEED* CMO. This again demonstrates the power and adaptability of the METIS approach, as the RSS plug-in can already serve as a standalone application without including it the context of the *MMS news application.*

Ontology-Driven Message Creation

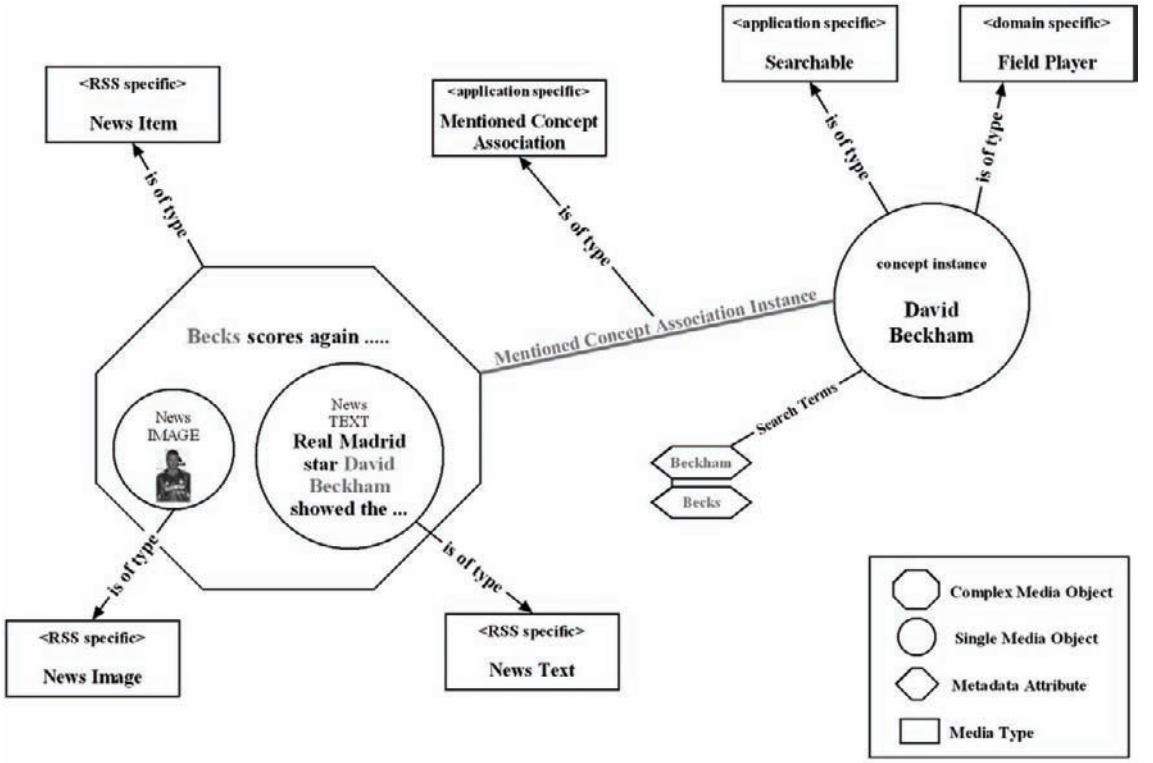
The *news application* plug-in provides core functionalities in the areas of ontology-based classification and discovery of specific media objects, as well as message creation from these search results.

The search terms provided by the knowledge base are used to identify textual occurrences of concept instances in *news ITEM CMOs.* We make the simplifying assumption that all other *media SMOs* included in this *news ITEM CMO* are also related to the discovered instance.

The news application plug-in uses this classification mechanism to relate concept instances to news items and their included media objects. A simple strategy based on regular expressions that searches all *news TEXT SMOs* for the occurrences of concept instance search terms defined in the knowledge base is currently implemented. This approach allows us to easily test the whole modular application on large datasets. Different search strategies can be utilised in this context and new ones can be added easily. For example, advanced full text analysis approaches could be employed in the application; this is a subject for our future research.

When a search term is found, a METIS association (of type *Mentioned Concept*) between the news text's *news ITEM CMO* and the concept instance SMO in the knowledge base is created, as depicted in Figure 8. This in turn fires a *new*

Figure 8. Concept mentioned association example



mentioned concept event that triggers the message creation mechanism.

Created messages are stored in a new container CMO of type *created message*. In most cases, news items contain only textual headlines; information and suitable media objects must be added in order to create a multimedia message for MMS delivery.

Once again, the domain-specific ontology provides valuable information about the relationships between a specific concept instance and other instances. As instances are bound to news items, the relationships can be derived for these news items as well. Media SMOs can thus be harvested from concept instances not bound to them, but bound to a closely related instance. Consider an example in which there are no images of the instance *David Beckham* available — in this case an image could be taken from an instance of *English National Team* as the latter is related to the former via an association of type *team member*. Only directly related concepts are taken into account, because we assume that the further apart two instances are, the more likely it is that unsuitable *media SMOs* will be chosen.

MMS Creation and Content Delivery

The purpose of the MMS plug-in is to assemble a MMS message from a message CMO (of type *created message*) and transmit it to subscribed users. This plug-in employs the METIS template mechanism to create suitable SMIL-based MMS slideshow presentations, including media objects supplied by the *created message*. The template includes placeholders that are dynamically replaced by the actual multimedia object instance data.

During the next step, the MMS message is packaged as a binary stream (because the MMS format does not allow any links to external media)

consisting of the actual media data referenced by the included SMOs and the generated SMIL file. General message attributes such as the receiver's phone number, the MMS title and subject, as supplied by the *created message* CMO, are also included in the header. That package is then sent to a MMSC, which continues by sending the MMS to the corresponding mobile device over a carrier's network.

This architecture has some specific advantages over other methods of sending MMS messages. First of all, the MMSC usually offers a mechanism for content adaptation and conversion according to a mobile phone's capabilities. This frees the METIS MMS plug-in from any consideration of the supplied media items in terms of conversion and adaptation to specific mobile devices. The second reason is that this design makes it possible to switch between MMSC implementations quite easily. Thus it is possible to adapt the *MMS news application* to any provider's or carrier's network architecture with a minimum amount of effort. Live environments that can send thousands of messages per second, compared to 2-4 messages in the testing environment, are therefore a future possibility.

CONCLUSION AND FUTURE WORK

Today, mobile multimedia applications provide customers with only limited means to define what kind of information they want to receive. Customers would prefer to receive information that reflects their specific personal interests, and this requires a mediation layer between the users and content that is capable of modelling complex semantic annotations and relationships. This will be a crucial characteristic of next-generation multimedia platforms.

In this chapter we have presented a prototype multimedia application that demonstrates this type of personalised content delivery. The development of the application was based on a custom multimedia middleware framework, METIS, which can be easily tailored to specific application needs. Our experience with the implementation demonstrated the rapid and modular development made possible by such a flexible middleware framework.

The example domain chosen to illustrate our approach is the Soccer World Cup. An ontology for personal news feeds from this domain was developed, and our experience indicates that similar ontologies and the corresponding knowledge bases for other domains can be created with very little effort. In any case, the application architecture is independent of the specific application domain.

The first module of our prototype application harvests media information from RSS feeds. As a result of the modular application architecture, one could easily integrate additional content sources (for example, encoded in NewsML) that are commercially available from many news agencies, in order to create a commercial application.

In the second module, harvested news items are classified according to the concepts given by the ontology. In our demonstrator application we employed simple text classification techniques, but again thanks to flexible system architecture, more advanced classification techniques can be developed without altering other system components. Future work will focus on more advanced methods of content classfication and on measuring the quality of aggregated media content.

In the final application module, multimedia news messages are composed and delivered to users, according to preferences specified during the subscription process. In the demonstrator we composed and delivered SMIL-based MMS messages to the mobile phones of registered users using a local MMSC. However, the integration with commercial MMSCs, enabling mass transmission of MMS messages, would require no additional implementation and minimal configuration effort.

In conclusion, we believe that the guiding principles for future mobile multimedia applications must be derived from personalised services

(i.e., "personalised content is king.") Through personalisation, such applications can provide the possibility for mobile service providers to improve customer retention and usage patterns through the created added value for the customer.

ACKNOWLEDGMENTS

This work was supported by the Austrian Federal Ministry of Economics and Labour.

REFERENCES

Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H., & Shadbolt, N. (2003). Automatic ontology-based knowledge extraction and tailored biography generation from the Web. *IEEE Intelligent Systems, 18*(1), 14–21.

Bulterman, D. C. A., & Rutledge, L. (2004). *SMIL 2.0. Interactive multimedia for Web and mobile devices series.* Heidelberg, Germany: X.media Publishing.

D'Alessio, D., Murray K., Schiaffino, R., & Kreshenbaum, A. (2000). Hierarchical text categorization. *Proceedings of the RIAO2000.*

Fernandez-Garcia, N., & Sanchez-Fernandez, L. (2004). Building an ontology for news applications. Poster Presentation. *Proceedings of the International Semantic Web Conference ISWC-2004*, Hiroshima, Japan.

Hammersley, B. (2003). *Content syndication with RSS.* Sebastopol, CA: O'Reilly.

Ho, S. Y., & Kwok, S. H. (2003). The attraction of personalized service for users in mobile commerce: An empirical study. *SIGecom Exchanges, 3*(4), 10-18.

IPTC. (2005). *International Press Telec Council (IPTC) Web site.* Retrieved May 15, 2005, from http://www.iptc.org

King, R., Popitsch, N., & Westermann, U. (2004). METIS—A flexible database solution for the management of multimedia assets. *Proceedings of the 10th International Workshop on Multimedia Information Systems (MIS 2004).*

Malladi, R., & Agrawal, D. P. (2002). Current and future applications of mobile and wireless networks. *Communications of the ACM, 45*(10), 144-146.

News. (2005). *NEWS (News Engine Web Services) Project Web Site*. Retrieved May 15, 2005, from http://www.news-project.com

NewsML. (2003). *NewsML Specification 1.2*. Retrieved May 15, 2005, from http://www.newsml.org/pages/spec_main.php

Nokia Technical Report. (2003). *How to create MMS services.* Retrieved May 15, 2005, from http://www.forum.nokia.com/main/1,,040,00.html?fsrParam=2-3-/main.html&fileID=3340

OMA. (2005). *Multimedia Messaging Service—Architecture overview.* Version 1.2. Open Mobile Alliance. Retrieved May 15, 2005, from http://www.openmobilealliance.org/release_program/docs/MMS/V1_2-20050301-A/OMA-MMS-ARCH-V1_2-20050301-A.pdf

Patel, C., Supekar, K., & Lee, Y. (2003). Ontogenie: Extracting ontology instances from WWW. *Proceedings of the ISWC2003.*

Prism. (2004). *Publishing Requirements for Industry Standard Metadata (PRISM) Specification 1.2.* IDEAlliance. Retrieved May 15, 2005, from http://www.prismstandard.org/specifications

Rao, B., & Minakakis, L. (2003). Evolution of mobile location-based services. *Communications of the ACM, 46*(12), 61-65.

Reuters. (2005). *Reuters NewsML Showcase Website.* Retrieved May 15, 2005, from http://about.reuters.com/newsml

Sakurai, S., & Suyama, A. (2005). An e-mail analysis method based on text mining techniques. *Applied Soft Computing.* In Press.

Sarker, S., & Wells, J. D. (2003). Understanding mobile handheld device use and adoption. *Communications of the ACM, 46*(12), 35-40.

Schober, J. P., Hermes, T., & Herzog, O. (2004). Content-based image retrieval by ontology-based object recognition. *Proceedings of the KI-2004 Workshop on Applications of Description Logics (ADL-2004).* Ulm, Germany.

Sony Ericsson Developers Guidelines. (2004). *Multimedia Messaging Service (MMS).* Retrieved May 15, 2005, from http://developer. sonyericsson. com/getDocument.do?docId= 65036

Stuckenschmidt, H., & van Harmelen, F. (2001). Ontology-based metadata generation from semi-structured information. *K-CAP 2001: Proceedings of the International Conference on Knowledge Capture* (pp. 163-170). New York.

Vlachos, P., & Vrechopoulos, A. (2004). Emerging customer trends towards mobile music services. *ICEC '04: Proceedings of the 6th International Conference on Electronic Commerce* (pp. 566-574). New York.

Vodafone. (2005). *Vodafone live! UK—MMS Sports Subscription Services.* Retrieved May 15, 2005, from http://www.vizzavi.co.uk/uk/sports-football.html

Wustemann, J. (2004). RSS: The latest feed. *Library Hi Tech, 22*(4), 404-413.

KEY TERMS

3G Mobile: Third generation mobile network, such as UMTS in Europe or CDMA2000 in the U.S. and Japan.

METIS: METIS is an intermedia middleware solution facilitating the exchange of data between diverse applications as well as the integration of diverse data sources, demantic searching and content adaptation for display on various publishing platforms.

MMS: Multimedia Messaging Service is a system used to transmit various kinds of multimedia messages and presentations over mobile networks.

NewsML: News Markup Language is an open XML-based electronic news standard used by major news providers to exchange news and stories and to facilitate the delivery of these to diverse receiving devices.

News Syndication: Is the process of making content available to a range of news subscribers free of charge or by licensing.

Ontology: A conceptual schema representing the knowledge of a certain domain of interest.

PRISM: Publishing Requirements for Industry Standard Metadata is a standard XML metadata volabulary for the publishing industry to facilitate syndicating, aggregating, and processing of content of any type.

Semantic Classification: Is the classification of multimedia objects and concepts and their interrelationships using semantic information provided by a domain schema (i.e., ontology).

SMIL: Synchronized Multimedia Integration Language is a XML-based language for integrating sets of multimedia objects into a multimedia presentation.

RSS: Really Simple Syndication (also Rich Site Summary and RDF Site Summary) is a XML-based syndication language that allows users to subscribe to news services provided by Web sites and Weblogs.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 232-250, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.11
Mobility Prediction for Multimedia Services

Damien Charlet
INRIA-Rocquencourt (ARLES Project), France

Frédéric Lassabe
University of Franche-Comté, France

Philippe Canalda
University of Franche-Comté, France

Pascal Chatonnay
University of Franche-Comté, France

François Spies
University of Franche-Comté, France

ABSTRACT

Advances in technology have enabled a proliferation of mobile devices and a broad spectrum of novel and out breaking solutions for new applications and services. In the present, more and more people and companies are demanding mobile access to multimedia services such as real-time rich media. Today, it is necessary to be able to predict adaptation behaviour which concerns and addresses not only the mobile usage or the infrastructure availability, but the service quality especially the continuity of service. Our chapter provides insight to new challenges of mobile multimedia services and applications: Wifi indoor positioning system adapted to heterogeneous building, static and learning mobility prediction, predictive handover policy for multimedia cache management, mobile multimedia guide (such as museum), and network scalability.

INTRODUCTION

The rapid deployment and growth of multimedia applications are increasing with the appearance

of new mobile services and new usages. Nowadays, taking advantage of the arrival of large bandwidth of wireless networks, it is becoming more feasible to stream numerous rich media flows towards mobile and terminal devices. However, some bottlenecks subsist when addressing, firstly, the heterogeneity of Wifi covered territories and secondly the intrinsic rich media constraints. We compare mobility to, first of all, a continuous move within a geographical space, and second a discrete space on a logical scale of the diffusion's network (from access point to access point).

This chapter deals with applications handling large size and continuous rich media communication (i.e., audio or video media). Continuous media require the installation of a specific infrastructure of diffusion guaranteeing the delivery periods. We are interested in mobiles implemented within a space provided with multiple access points, with a more or less homogeneous space cover. In such context, it is important the infrastructure reacts rapidly, or use preventive measures during the changes of access point.

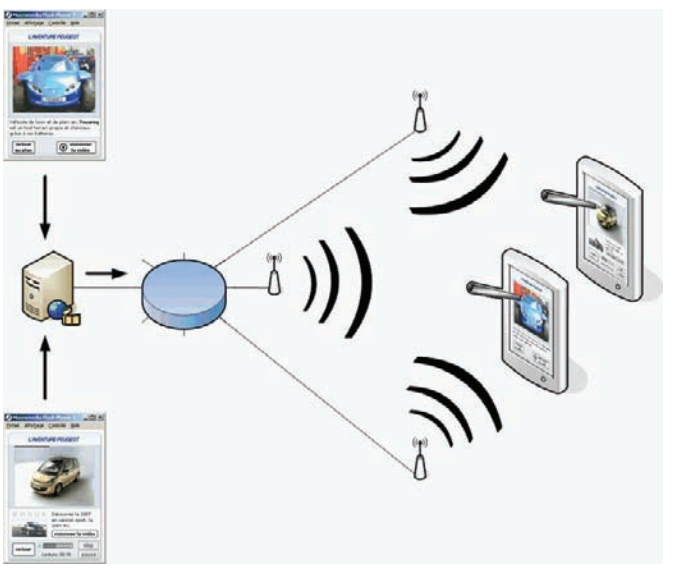
In this chapter, we do not consider the dynamic flow adaptation but rather, we consider already optimized flows dedicated to mobile devices. Thus, whatever the device nature is (pda, tablet pc, etc.), we assume there is a suitable flow adapted to each target. The reader interested in flow adaptation may refer to (Bourgeois, Mory, & Spies, 2003).

To illustrate our purpose, we use GUINUMO, a mobile numerical guide. Such guide demonstrates the accuracy and pertinence of retrieving and making use of both the visual or audio information, and the localization of the pervasive device, during the time-visit of scenarized museums. Within this framework, the media are suited to fit the specific device.

In the sequel, first of all we present the techniques of localization of the devices connected by hertzian way. We further investigate the trilateration technique and evaluate the efficiency of various methods according to several conditions of implementation.

In order to set up a preventive treatment of mobility we show that it is necessary to determine, at least statistically, the future position of a mobile. Then, we describe the methods allowing predicting, in the short run, the position of a mobile. We detail how, thanks to methods of training, it is possible to refine this prediction.

Figure 1. Synthetic schema of GUINUMO's platform



In the second part of this chapter, we introduce the concept of cache, as a necessary element in the chain of continuous media diffusion. Caches make it possible to ensure the continuity and the extensibility of the diffusion's infrastructure. We start off by describing the standard methods of managements and co-operation of the caches for continuous media. We proceed by explaining the mechanisms required to manage the change of access point: *handoff*. Then we detail how preventive methods allow optimizing the continuity of flows diffusion. We also present how to integrate these mechanisms in a platform of diffusion and reception (GUINUMO). We describe use-cases of this device. Finally, we conclude with future trends about preventive treatment of mobility. We specify how the coming standards will allow optimizing the handoff and positioning determination mechanisms.

CONSIDERING MOBILITY

Positioning

Positioning is required to provide location-related services. By positioning we mean determining a mobile's geographical coordinates.

Major Positioning Techniques

Global positioning system (GPS). The GPS (Alouini, 1996) is the system allowing locating an apparatus on the surface of the planet or in the atmosphere, using reference satellites whose positions are known. This system has been essential for several years. Thanks to preliminary synchronization of the satellites with the GPS receiver, the latter can compute, based on the time of course between the satellites and the receiver, the distance separating it from each satellite. Trilateration (Fang, 1986) means the intersection of the spheres determined by the distances between the mobile device and each satellite and the receiver's position is calculated. The precision is about ±5 meters for civil applications. The main drawback is that it does not function well enough indoors, in city centres, or in raised or leafy environments.

GSM terminal positioning, standing for global system for mobile communication (GSM), is the main mobile telephony standard in use. This standard has given birth to three positioning methods. For each method, setup and accuracy vary. The methods are the *positioning by cell*, the *computation of the distance according to the signal strength,* and the *distance computation by time difference.*

The positioning by cell is basic. Interrogating the BTS (base transmitting station) is sufficient to identify to which one the mobile is connected. The mobile phone is located in the cover area of the said BTS. This method has a variable accuracy, from 100 meters to several miles (depending on the cell size, itself depending on the mobile phones density). In cases where a service like calling the mobile phone is provided, such accuracy is enough.

The positioning by measuring the signal strength is simple too. Knowing the radio wave weakening according to the distance between the transmitter and the receiver, the distances towards the BTS can be extrapolated by measuring the signal strength. With the distances towards the BTS (which positions are known), the trilateration, exposed in the paragraph concerning GPS, gives the position of the mobile phone. The accuracy obtained is between 50 and 500 meters.

Enhanced observed time difference (EOTD) is a method to estimate the time for the signal to go from the mobile phone to the BTS. This technique is usually employed to adapt the transmission timing of the mobile phones according to their distance to the BTS, thus allowing the scheduling of the packets in the time slots. Without this method, the packets would come to collision near the BTS. The adaptation of the transmission timing is called Timing Advance and requires

the synchronization of the BTS and the mobile phones. The distance computation with the time difference is based on EOTD to compute the distance between the BTS and the mobile phone. Knowing the time used by the signal to go from the mobile to the BTS and the wave speed, we can compute the distance. Then, trilateration gives the position of the mobile phone.

Wifi positioning techniques can be classified into two main categories, the one based on signal *strength cartography*, and the other which determines a relation between signal strength and distance. That makes the location computation possible using trilateration.

Within the RADAR system (Bahl & Padmanabhan, 2000), the mobile terminal positioning uses a signal strength map of the covered area. The geographic coordinates, the signal strength measurements and the mobile orientation are stored in a database. The signal strength map can either be constituted by computation or by physical measurements. The signal strength measurement from each access point is compared with the reference points stored in the database. The cartography-based positioning technique has a 2-to-3-meter precision.

Wang, Jia, and Lee (2003) present a *positioning technique based on a radio wave propagation model*. This model aims at expressing the mathematical relation between the distance from transmitter to receiver and the signal strength. The mathematical expression is obtained by polynomial regression of the third degree. The advantage of this technique is the speed of positioning. However, there is a main drawback. A lot of data are required for the regression to be accurate, which involves a high cost in measurement time. On top of that, it is possible to be confronted with singularities in the buildings where the positioning technique is implemented.

The white paper of Interlink Networks (2002) deals with security issues. Its first objective is to locate rogue mobile terminals and access points which try to infiltrate a network through its wireless part. The authors take signal strength measurements at many locations of many buildings. The results of these measurements are used to establish a radio wave propagation model. This model is based on the Friis relation. The Friis relation expresses the signal strength in function of distance, in a free space environment. The Friis-based model is adapted to fit the conditions of implementation. The precision observed is close to 2 meters. The main advantage of this technique is its setup speed. However, some singular geographic points were observed where the precision was worse than 8 meters. The main drawback of this technique is the unique exponent used in the Friis equation.

The best precision is obtained by the signal strength cartography-based technique. However, it uses lots of resources and computing time to use a signal strength map. A long setup time is also required, and it has no reactivity when topological changes occur. These drawbacks partially affect the polynomial regression-based technique because of the need for data in order to obtain the polynomial expression of the distance. Although its precision is less accurate than that of the previous technique, the technique based on an alternative to the Friis equation is very quick to setup and use. Thus, it is well adapted to topological changes.

Singular points are intrinsic to the topological heterogeneity. Buildings are composed of obstacles which interfere with radio wave propagation. The obstacles can be of various natures (Larnon, 1998) and their layout can be irregular. When facing such unfavourable cases, the signal strength cartography shows better results because it fits the building, whereas the propagation model-based techniques consider the topology uniformly.

GeoMovie

The Interlink Networks (2002) approach is chosen to implement our positioning system. It has indeed the advantages of speed and simplicity. It

is interesting with mobile terminals which have little computation power. In Lassabe, Baala, Canalda, Chatonnay, and Spies (2005), we explain the drawbacks of a uniform computation in order to determine the distance according to the signal strength. We first describe the common sources of radio wave distortion and their predominance within a heterogeneous environment. Second, we highlight the radio wave distortion indoor environments with the help of our experiments and we test the model of Interlink Networks, to reveal its limits in a heterogeneous environment.

The Friis equation:

$$\frac{P_R}{P_T} = G_R G_T (\frac{\lambda}{4\pi d})^2$$

where P_R and P_T are respectively the power available at the receiving antenna and the power supplied to the source antenna;

- G_R and G_T are respectively the receiver antenna gain and the transmitter antenna gain
- λ is the carrier wavelength
- d is the transmitter-receiver distance

The Friis equation expresses the signal strength loss in function of the distance *d*. The radio wave absorption by obstacles is similar to the free-space loss but it is generally greater.

We use a reciprocal expression to the Friis equation to determine a value replacing the square of the distance adapted to the environment where the positioning will be achieved. Sample measurements are used to compute the value adapted, which is different for each access point. Then, we use the expression of the distance according to the signal strength, taking into account the new coefficient in the such-modified Friis equation. Trilateration for itself is achieved by an algorithm trying to minimize the distance between the circles centred on the access points and whose radius are the distances towards the mobile terminal (see Figure 2).

The radio waves are affected by the presence of topological components altering the radio waves trajectory and therefore modify the signal strength. The phenomenon we are more likely to observe is wave reflection. The most common sources of wave trajectory distortion are metal equipment that induces huge signal reflections, preventing it from reaching areas theoretically within range. Devices functioning at frequencies close to Wifi frequencies also distort the signal by covering it with great noise.

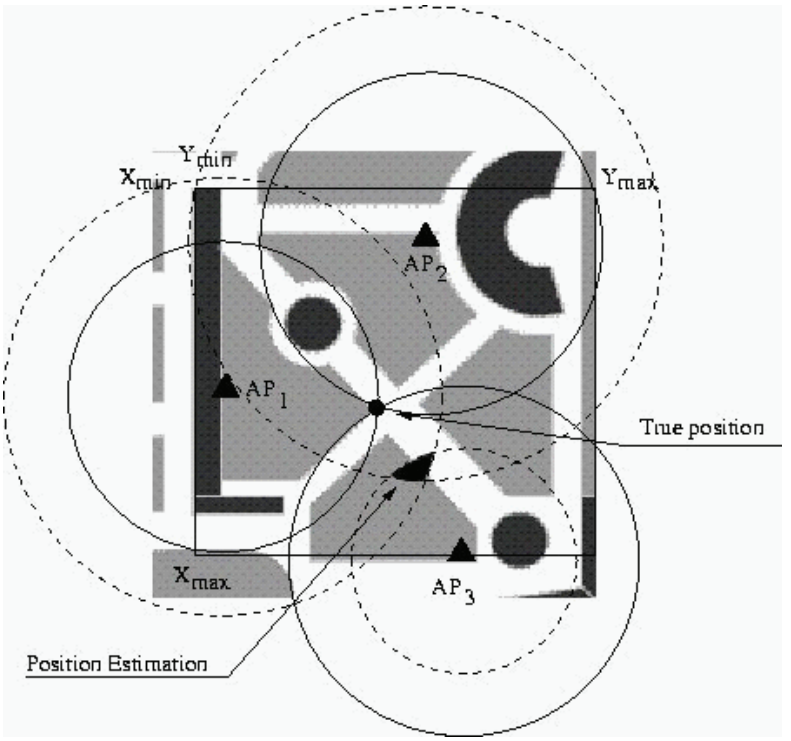
Predictive Positioning

Knowing a mobile terminal position is not sufficient to ensure service continuity. Mobility prediction is required to do so. Predicting the mobile terminal movements requires modelling the trajectories of the mobile terminals. We present two mobility models. The first one models the *trajectories by the movement vectors* of the mobile terminals. It is a static method in which a simple trajectory equation is used. The second one is based on the *learning of the mobile terminal movements*. It is a dynamic method which can be updated when the users change their movements habits. In particular, we present the hidden Markov model (HMM) (Rabiner, 1989). These learning models can be extended by taking into account some mobility patterns.

Trajectory Calculation

In a project to increase routing efficiency in *ad hoc* networks (Lee, Su, & Gerla, 2000), mobility prediction is used. A mobile terminal trajectory at a given time is modelled by the direction and the coordinates of the speed vector of the mobile terminal trajectory. These data are sufficient to anticipate the future position of a mobile terminal. They also allow estimating the remaining connection time between two mobile terminals.

Figure 2. Geolocation based on trilateration technic



Simulations based on this model show a growth of the routing efficiency. However, the experiment hypotheses are not always realized in real conditions. The hypotheses are a positioning of the mobile terminal without error and the uniformity of its trajectory. If the mobile terminal follows a complex course, the routing loses its efficiency.

Learning from the Movements

Another mobility management method is based on the learning of the mobile terminal movements. The learning can be based on several models such as Markov model or automata. Learning the moves allow to compute transition probabilities from state to state. It is interesting because the states can be physical areas or logical states such as the dependency from a multimedia cache or the connection to a base station (in GSM, or access point in Wifi).

Use of Hidden Markov Model

In the article from the University of Liège (François, Leduc, & Martin, 2003), a mobility prediction model for mobile networks is presented. It is based on the hidden Markov model. The transition probabilities between the states are determined by a learning system. During the learning, the mobile terminals regularly send their position to the base stations. The base stations save the transitions from a state to another and compute the transition probabilities.

Simulations without positioning errors are used. The model accuracy is sufficient to predict the mobile terminal next movements. In further work (François, Leduc, & Martin, 2004), an error is added to the mobile terminal positioning. The results are less accurate but still good. When using the first two results, the prediction is 75% right with a 15% noise on the signal strength and

five observations to compute the next state. These conditions are the worst tested.

Mobility Patterns

Mobility patterns are built from logs of several visits. They are used to identify classes of equivalence. Based on such identified classes of equivalence, it is possible to characterise some visitors' behaviour by attaching to visitors (Chardonnel & Van Der Knaap, 2002), and during their visit, the classes matched with behaviour logs, for example. Then, when a user is attached to one (or more) classe(s), it becomes easier to be accurate when predicting where they are going to be located and which media he is going to ask for.

Mobility patterns matching is made on the fly during visits. Thus, as the behaviour of a visitor may change during their visit, class matching must provide a mechanism for identifying new behaviour and adapt any predictions which concern either the next geolocation, or the next interaction with the mobile terminal, or else solicitation of new media. Mobility patterns are build wether online or offline, using the log file of the system.

STREAMING FOR MANY MOBILE USERS

We have seen, in the previous section, that it is possible to locate a mobile terminal in the physical space. It is achieved with accuracy depending on several criteria. The knowledge of the position also involves a possibility to predict the close future moves of the mobile terminal. The best methods are actually based on movements learning. Both the positioning and the mobility prediction allow the service continuity.

Streaming video to many mobile users brings problems which may be categorized in two classes. Firstly, the huge amount of data which has to be sent, blocks the network and the servers, and decreases the interactivity. And secondly, the tight real time requirements of video streaming cannot cope with the latency induced by the distance of clients moving away from the source.

The first class of problems may be solved by using a cache (Cao & Irani, 1997). A cache may be seen as an empty server at time 0. It may transmit the objects it possesses, like a server. Its particularity resides in the fact that it may request objects from other servers or caches and store them temporarily to serve them later. Caches are placed near the clients to reduce the length of the paths and therefore enhance the interactivity and avoid bottlenecks. Moreover, to cover wide areas and serve many users, several caches may cooperate in a set of distributed and cooperating caches.

Caches are managed by several policies: the role of the insertion policy is to decide when and which new documents should be stored on the cache, the removal policy is dedicated to the cleaning of the storage space, the admission policy considers if we can accept a new client, and the aim of the sibling (or cooperation) policy is to take care of the content of the neighbouring caches in a distributed system. Many examples of such policies may be found in the literature (Balamash & Krunz, 2004), some of them are dedicated to specific content such as video (Podlipnig & Boszormenyi, 2002; Rejaie & Kangasharju, 2001). However, none of these policies cope with the problems related to the mobility of the clients.

Mobility in Video Caches

Taking into account the mobility of the users brings new constraints to the field of caches. These not only store and serve the data according to the requests they receive any more, but they must also manage the mobility of the clients. Indeed, the usual operation of the caches is based on the observation of the requests they receive. In a traditional scheme, the topology of the served clients remains fix, and the uses similar along the time. It is thus enough for the caches to adapt their content

and their operation to be effectively useful. In a mobile context, as the clients are regularly moving from one zone to another, topology is always fluctuating. A client used by a cache at time T may be out of reach at the next time step.

The integration of the mobility in the caches is a recent problem, and few solutions are exposed in the literature. Two families of solutions may be distinguished: an optimized co-operation of caches, which may imply a change of cache and a context switch where the caches become mobiles in order to follow displacements of the clients.

In Hadjiefthymiades and Merakos (2001), the authors place themselves within the framework of a cellular network in which mobile clients consult Web sites. In order to optimize the navigation of the clients as well as possible, the authors propose to insert a proxy-cache for each client within the system. This one is placed close to the user, at the level of the base stations, and stores a small set of Web pages and pictures useful for the navigation of the client. Its contents being limited, it can, at lower cost, move to follow the displacements of the client. This technique is interesting but requires that the whole end-side material (access point or base station) should be equipped with sufficient capabilities of calculation and storage. Moreover, the migration of a cache can raise the cost of the network if the stored documents are bulky, as within a video framework.

Some projects descend the level of the cache since it is placed on the client himself. Thus in Cohen, Herscovici, Petruschka, Maarek, and Soffer (2002), the authors use a cache placed on the mobile device, which is in charge of fetching a set of pages related to the navigation in progress, as soon as possible. The authors of Sailhan and Issarny (2003) present a system of caches for mobile clients functioning in *ad hoc* mode. This system is particularly useful for the users exploiting their wandering peripherals in zones where many other peers are. The example given is that of a museum offering a virtual visit thanks to personal assistants. In this architecture, the PDAs obtain their information on a server via access points disseminated through the museum. It may happen that no access point is within the range of a client. If a networked resource is to be reached, it is possible for this client to connect in *ad hoc* mode to terminals available in its neighbourhood, and to launch a strategy of exchange of cached information. This strategy is very interesting since the caches are placed on the same level as the clients, and the cooperative mechanism makes it possible to optimize the use of the bandwidth. However, it requires the use of peripherals having a significant storage capacity, and those must incorporate the cache program, which is not always possible.

Within the framework of video streaming, the mobility of the cache is hardly possible. Indeed, the volume of transmitted data being very significant, a cache migrating from one place to the other implies a very high cost of utilization. The mobility of the users have thus to be treated by optimized policies of inter-cache co-operations.

Handoff Policy to Tackle Mobility Between Caches

An optimized policy of cooperation of caches for the mobility of the users has a main function: to envisage the change of cache of the users because of their mobility. It is indeed useless to continue diffusion between a cache and a client which would have moved away too far from the former, when another more optimal cache could take over.

When a set of distributed caches is deployed, a paving of the territory in zones of diffusion can be imagined. The policy of cooperation must then supervise displacements of the client in order to detect a change of zone. This roaming from one zone to another may be compared to the change of terminal in mobile telephony or the wireless networks. One will then speak of handoff.

The first basic policy consists in carrying out no action whatsoever in any case. This corresponds to what currently occurs when no

policy is implemented. The problem that arises is the lack of reactivity in the event of a handoff. Indeed, when a user looking at a film changes zones, the cache, which must take over, does not have the sequence requested and was not advised of a possible arrival of the request. The diffusion must then stop while the part being streamed is fetched in the new cache.

The second basic policy (called "broadcast") consists in contacting all the neighbouring caches to advise them to fetch the currently streamed sequence. The defect of reactivity is thus smoothed out, each cache keeping the sequence in the event a user should arrive. However, this solution is not satisfactory because it quickly causes a clogging on the network between the caches and it uses disk space of the caches for sequences which have few chances to be used.

In order to optimize the transition when a client moves, thus causing a change of zones, the caches must cooperate by exchanging information with their "neighbours" and by prefetching part or whole of one or several sequences. An optimized management should thus define a set of caches sufficient at the same time to ensure a strong probability of hit while minimizing the disk space and the lost bandwidth.

Use of the Prediction of Position in Video Caches

To succeed in treating the handoff (i.e., the change of cache in a short time), compatible with the temporal constraints of the diffusion, and also by mobilizing only the necessary resources in order not to disturb too much the diffusion towards the other users, the system cannot be only reactive. The system of caches needs to anticipate this possibility and the potentially concerned caches should be all set to stream before the user changes zones. In addition, only a minimal number of caches should be included in this preventive measure in order not to put the scalability of the system in danger.

We propose to prefetch, in a reduced number of caches, the continuation of the sequence in the course of visualization of the users who are likely to change zones. With this purpose in mind, our technique is based on an observation of the mobility of the users, on a prediction of their short-term position and on the adequacy of each cache to serve the zones having to receive new users.

Our handoff policy is based on external information to anticipate the necessity of prefetching a cache. This information is of two types: the probability for a zone to receive a client in an immediate future, and the adequacy of the connections between the caches and the zones of reception. From this information we calculate an indicator evaluating the need to fetch the sequences.

When a client is in a given zone, we need to get the set of the probabilities of presence of the client in all nearby zones, at the next moment. Our policy is based on two view points resulting from the observation of former displacements: the observation, on the one hand, of the former moves carried out by the whole set of clients and,

Figure 3. Computation of the relevant pre fetching vector

$$\mathcal{M}_{(adequacy)} \times \vec{V_p} = \vec{V_e}$$

$$\begin{pmatrix} Cache_0 Zone_0 & \dots & C_0 Z_z \\ \vdots & \ddots & \vdots \\ C_c Z_0 & \dots & C_c Z_z \end{pmatrix} \times \begin{vmatrix} P_0 \\ \vdots \\ P_z \end{vmatrix} = \begin{vmatrix} e_0 \\ \vdots \\ e_c \end{vmatrix}$$

on the other hand, of those of the particular client. In order to refine the behaviour of the system, these two approaches are balanced thanks to an exponential average. We thus obtain a probability of presence for each close zone. At the end of this calculation, we present this set of probabilities in the form of a vector called *Vp* which dimension is the number of close zones plus 1 for the zone in progress. This last value, noted P0, represents the probability that the user will remain in the zone.

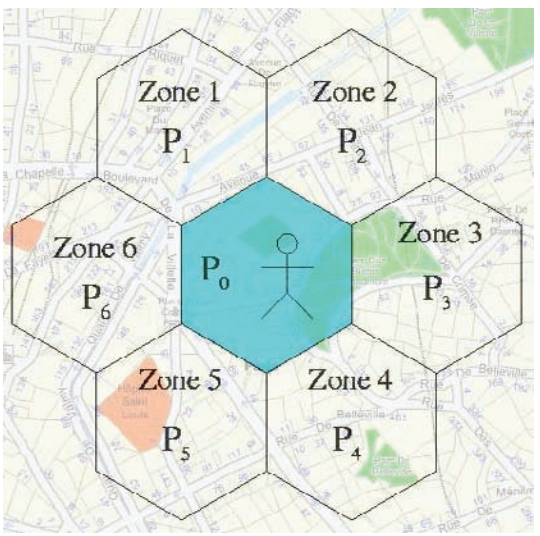
The adequacy between a cache and a zone is an indicator which quantifies the effectiveness with which a cache can stream a sequence towards a zone. The cache-zone adequacy does not take the availability of a particular sequence in a cache into account, but qualifies the quality of the connection. A simple example of quantifier is the average Round Trip Time between a cache and a zone. It is however necessary to integrate the variations of quality in the course of time, and to recompute this value regularly. These factors of adequacy are represented in the shape of a matrix binding each cache potentially useful for the diffusion to each zone, built in a distributed way thanks to exchanges between the caches.

That still remains to quantify the relevance of caches fetching. The columns of the matrix, representing the zones, are sorted in such way as to correspond to the lines of the vector *Vp*. Then, the vector *Ve* associating to each cache a quantity, which is the sum of the products of the cache-zone adequacies by the probability of presence of the user in the zone, is defined by the product of the matrix by the vector *Vp*. The resulting values of the vector are in a range from 0 to 1. Indeed, the vector of probability of presence is a probability distribution whose sum is equal to 1. The coefficients of the matrix of adequacy are, by construction, all ranging between 0 and 1. Thus, the sum of the products of the ones by the others also lies between 0 and 1. The larger the values of the resulting vector are, the most suited for streaming to a zone having a large probability of receiving a client a cache is.

Since we have an evaluation of the interest of fetching for each cache, we still have to define the level needed to load this cache. This threshold can be modulated in order to define the level of aggressiveness of the policy. For example, in order take a multi-layers stream into account (a basic stream and a complementary stream), a first variable HS = P0 * threshold will give the value of swing beyond which the entirety of flow will be charged, and a variable $Sb = a * HS$ makes it possible to define a range within which only the basic stream will be transferred. Lastly, if the evaluation is lower than Sb, nothing will be charged.

This method makes it possible to improve the fluidity of the diffusion towards mobile clients by being based on a short-term prediction of their future positions. Thanks to the dynamic construction of a matrix of adequacy between a set of caches and a set of zones of diffusion, and with the use of the probabilities of presence, we deduce an algorithm revealing the state in which the system will have to be at moment *T+1* will have to be. Analysing this algorithm will allow us

Figure 4. Territory partitioning in streaming zones



to set up a policy of total or partial prefetching of the sequences in the neighbouring caches.

GUIDE NUMÉRIQUE MOBILE: GUINUMO

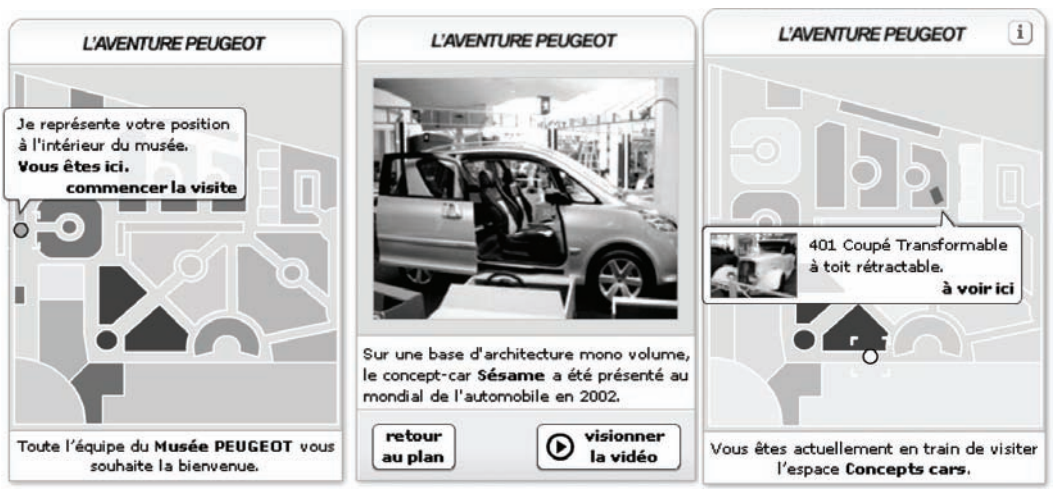
We have a demonstrator of the techniques we presented above. This product, called GuiNuMo, allows the assisted visit of the "Museum of the Peugeot Adventure." Our tool allows the geolocation of the visitor within the museum and the on-demand streaming of audio-visual content related to the environment they will come across. In Figure 4 you can see three GuiNuMo screenshots on a PDA. The picture on the left is that of the beginning of the visit; one discovers the plan of the museum and the position of the visitor, near to the entry. On the map of the museum the various zones of diffusion are shown. They are associated to themed elements of the museum (concept car, competition...). The screenshot in the middle presents the GuiNuMo interface when the user enters the zone called « concept car ». The presentation of a specific vehicle is offered and it is possible to play a video. The video, if requested, will be streamed from the cache dedicated to this zone.

During the whole visit, the position of the peripheral is computed in real time, and the available content is provided in the form of play lists that the user can choose to visualize or not. When it is possible, GuiNuMo suggests the visitor to go to a different area of the museum in order to discover objects in relation to the media he has just consulted. For example, at the end of a video depicting the "Peugeot 306 CC," GuiNuMo will suggest going to the part devoted to the "Peugeot 401 coupé transformable" which was, in 1933, the first car integrating a retractable roof. See screenshot 3 for a graphic representation.

All the actions of the visitor are recorded by the system in order to be able to analyze the visit and to feed the training algorithm. Moreover the visit can be studied by the staff of the museum to deduce some potential transformations. Indeed by analyzing the logs of the visits of multiple visitors, it is possible to note that few people visit certain sites in the museum or that a particular zone usually holds greater attention. Thus, the team of the museum may choose to rearrange spaces or to make recommendations in order to enlight under exploited spaces.

The training algorithm makes it possible to determine, statistically, the probability of visualization of each media according to the actions

Figure 5. Three screen shots of GUINUMO application interface



carried out previously. Consequently, it is possible to anticipate which media will probably be visualized soon. It is thus possible to prepare the system to allow a better reactivity.

Right now, the prediction is carried out as a whole. We are currently working on the identification of the profile of the users in order to obtain a prediction of better quality. Indeed, each visitor does not visualize all the available media, but operates a themed selection set. One can make a difference between the visitors interested in technology and those interested in the history of the Peugeot Company. By analyzing the first documents requested by the user, it is possible to determine to which group of users they belong and to subsequently direct the prediction and the recommendations according to the actions carried out previously by other users of the same class of equivalence.

CONCLUSION AND FUTURE TRENDS

In this chapter, we described the interest of the predictive methods during the broadcast of multimedia flows towards mobile terminals. These methods are integrated in location services as predictions and in the architecture as caches. Caches allow the scalability of the system and the storage of data to send to the client into the future probabilistic areas.

The location predictions based on the knowledge, allows the anticipation of the possible or probable areas of location. This method is included in the cache loading strategies as well as in the establishment and continuity of the terminal connection. The union of these two mechanisms allows a fast and accurate management of the access point migration when a terminal moves from one zone to another.

The intercache communication methods, called "cache sibling," are well-known algorithms. They make it possible to manage the content deletion in the distributed caches. They require to be adapted, to take into account the specificity of the video caches. Indeed, a video cache does not maintain strong coherence. It has to communicate with its "neighbours" to facilitate the continuity of the sequence broadcast.

The localization principle used in optical or hertzian ways is rather old. It is well adapted to outdoor spaces (e.g., on seas). It is possible to use two great classes of algorithms: the triangulation or the trilateration. For clients, in Wifi-like networks, we do not have access to angular information; it is thus the trilateration which is used. In the case of the indoor localization, corrections must be made to the coefficient of the Friis equation in order to take the absorption of the wave by the building elements into account.

The prediction methods of the mobile location can be based on two types of information: a topological knowledge of the area, allowing the identification of relevant trajectories, and experiences of the previous movements deduced from a position list, the future most probable positions. The experience can be limited to previous moves of the client which receives the flow or can combine information resulting from the movements of all the users having crossed the considered area. A relevant method consists in mixing general and individual information.

The GuiNuMo1 application which was presented in this chapter to illustrate our methods is installed in a museum for demonstration purposes. This experimental platform enables us to work on real scenarios and to observe the reactivity of the system. The collected information tends to ascertain the hypothesis that mobile multimedia applications are more and more accepted by users, and drive us to consider implementing new features.

Using caches as the broadcast infrastructure of multimedia flow is a significant point. We are now thinking of installing caches into the client device and of sharing content of these caches in an *ad hoc*, peer-to-peer network. This method should allow under-dimensioning, possibly ever to remove the cache from the broadcast infrastruc-

ture, in order to rely strictly the mobile devices. In case of numerous clients, the requirement of caches is important. The available space and the density of redundant data are linked to the number of clients which will allow scalability.

The vertical handoff is a research topic currently studied by a great number of international teams. It is the opportunity to switch from a network to another, keeping a connection continuity of the current applications. Basically, the IP protocol does not offer such service, and mobileIP allows only a discrete mobility, which means without any continuity on the current connections. To achieve a goal of continuity, it is necessary to adapt or modify the TCP/IP stack of the terminal and the servers. It is also possible to modify the TCP/IP stack of a network device near the client as the access-point, in place of the servers, in order to manage this problem. The vertical handoff must integrate a policy of network selection. In fact, it is important to select the right network, which means the network with the best bandwidth. But, with this strategy, we do not have the opportunity to choose the best available throughput. In order to do so, it is necessary to dynamically choose the best network, causing additional handoff in case of network overloads. Moreover, it will be possible to connect few networks at the same time in order to have a sum of download or upload throughputs using a multi path strategy.

The step after the vertical handoff will study the management of network failures mainly due to white spots using strategies of mobile perfecting. This technique will allow to load in advance some documents during the connected times in order to deliver them latter even if a failure network occurs during the request of view.

REFERENCES

Alouini, M. (1996). Global positioning system: An overview. *Tunisian Scientific Magazine, 10*(1), 49-51.

Bahl, P., & Padmanabhan, V. N. (2000). RADAR: An in-building RF-based user location and tracking system. *Proceedings of INFOCOM, 2*, 775-784.

Balamash, A., & Krunz, M. (2004). An overview of Web caching replacement algorithms. *IEEE Communications Surveys and Tutorials, 6*(2), 44-56.

Bourgeois, J., Mory, E., & Spies, F. (2003, November). Video transmission adaptation on mobile devices. *Journal of Systems Architecture, 49*(10-11), 475-484.

Cao, P., & Irani, S. (1997, May). *Cost-aware WWW proxy caching algorithm* (Tech. Rep. No. CS-TR-1997-1343). Madison, WI: University of Wisconsin.

Chardonnel, S., & Van Der Knaap, W. G. M. (2002). Managing tourist time-space movements in recreational areas: A comparison between two areas with the same analysis methodology for a protected nature park in the French Alps and the Dutch National Park "De Hoge Veluwe. *Revue de Géographie Alpine, Tome, 90*(1), 37-48.

Cohen, D., Herscovici, M., Petruschka, Y., Maarek, Y. S., & Soffer, A. (2002). Personalized pocket directories for mobile devices. *Proceedings of the 11th International Conference on World Wide Web* (pp. 627-638). ACM Press.

Fang, B. T. (1986). Trilateration and extension to global positioning system navigation. *Journal of Guidance, Control, and Dynamics, 9*(6), 715-717.

François, J. M., Leduc, G., & Martin, S. (2003). Evaluation d'une méthode de prédiction des déplacements de terminaux dans les réseaux mobiles, Réseaux mobiles et ad hoc, qualité de service, test et validation, ingénierie du traffic. *Special issue of Hermès Lavoisier* (pp. 189-202).

François, J. M., Leduc, G., & Martin, S. (2004). Learning movement patterns in mobile networks:

A generic method. *European Wireless 2004* (pp. 128-134).

Hadjiefthymiades, S., & Merakos, L. (2001, May). Using proxy cache relocation to accelerate Web browsing in wireless/mobile communications. In *WWW'10 Conference* (pp. 26-35).

Interlink Networks, Inc. (2002). *A practical approach to identifying and tracking Unauthorized 802.11 Cards and Access Points.* Technical Report. Retrieved from http://www.interlinknetworks.com/graphics/news/wireless_detection_and*_*tracking.pdf

Lassabe, F., Baala, O., Canalda, P., Chatonnay, P., & Spies, F. (2005). A Friis-based calibrated model for WiFi terminals positioning. *Proceedings of IEEE WoWMoM* (pp. 382-387).

Lee, S. J., Su, W., & Gerla, M. (2000). Mobility prediction in wireless networks. *Proceedings of IEEE ICCCN* (pp. 22-25).

McLarnon, B. (1998). *VHF/UHF/microwave radio propagation: A primer for digital experimenters.* TAPR's Spread Spectrum Update, Tucson Amateur Packet Radio Corporation. Retrieved from http://www. raveontech.com/Application-Notes/Primer.pdf

Podlipnig, S. & Boszormenyi, L. (2002). Replacement strategies for quality based video caching. *IEEE ICME'02, 2,* 49-52.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *IEEE, 77*(2), 257-286.

Rejaie, R., & Kangasharju, J. (2001). Mocha: A quality adaptative multimedia proxy cache for Internet streaming. *ACM NOSSDAV'01* (pp. 3-10).

Sailhan, F., & Issarny, V. (2003). Cooperative caching in *ad hoc* networks. *The 4th International Conference on Mobile Data Management* (Vol. LNCS 2574) (pp. 13-28).

Wang, Y., Jia, X., & Lee, H. K. (2003). An indoors wireless positioning system based on wireless local area network infrastructure. *The 6th International Symp. on Satellite Navigation Technology Including Mobile Positioning & Location Services* [CD-ROM, paper 54].

KEY TERMS

Ad Hoc Mode: Every client can talk to each other on a peer-to-peer basis.

Admission Policy: Algorithm used when a new client wants to fetch data from a cache to decide if the cache has sufficient capabilities left to serve him.

Cache: A cache gathers the functions of a server and of client. It takes place between them and can store and deliver popular documents. Being near the client, it helps resolve the problems of bottlenecks and increase reactivity.

Cooperation Policy: Used between distributed caches to cooperate and share data.

Handoff: Name of the mechanism which takes place when a user is roaming.

Insertion Policy: Algorithm of caches, decides which documents should be stored.

Mobile Network: Network in which part or all of the components are mobile.

Mobile Terminal: Every apparatus light-enough to be humanly transported and with embedded computation power, like laptops, PDA, new generation mobile phones.

Mobility: The action to move. We are interested in particular in the logical mobility (change of network, BTS, etc.) triggered by the geographical move (the action of changing physical coordinates in space).

Prefetch: Inserting documents into a cache in the hope that they are going to be requested in a near future to reduce start latency for the user. Contrarily to normal insertion, it is not triggered by clients.

Removal Policy: Management algorithm of caches, decides which stored documents should be deleted to make room for new and more popular documents.

Roaming: Action of a human moving from one zone to another.

Service Continuity: Property of a service over a mobile network. When continue, a service is not interrupted by changes in its logical position (change of AP / BTS). For example in the GSM standard, as long as you stay in covered areas, phone conversations are not interrupted when you change your BTS.

Sibling: Exchange of data between two caches.

Signal Strength: It is the power of the signal measured.

Start Latency: Time elapsed between the moment where a user requests of a document and the time it is displayed on its peripheral.

Streaming: Technique of transfer in a continuous flow to allow the display of the media while downloading.

Video Cache: A cache with specific policies, optimized for the delivery of video data.

ENDNOTE

[1] GuiNuMo is a project funded by: EU, french ministry of research, Franche-Comté Council and CAPM.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 491-506, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.12
Multicast of Multimedia Data

Christos Bouras
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

Apostolos Gkamas
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

Dimitris Primpas
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

Kostas Stamos
Research Academic Computer Technology Institute, Greece
University of Patras, Greece

INTRODUCTION

The heterogeneous network environment that Internet provides to real time applications as well as the lack of sufficient QoS (quality of service) guarantees, many times forces applications to embody adaptation schemes in order to work efficiently. In addition, any application that transmits data over the Internet should have a friendly behaviour toward the other flows that coexist in today's Internet and especially toward the *TCP* flows that comprise the majority of flows. We define as TCP friendly flow, a flow that consumes no more bandwidth than a TCP connection, which is traversing the same path with that flow (Pandhye, Kurose, Towsley, & Koodli, 1999).

During the multicast transmission over the Internet, several aspects need to be considered:

- **Transmission rate adaptation:** The sender must adapt the transmission rate based on the current network conditions.
- **TCP friendliness:** During the multicast transmission over the Internet, the multicasts flows must be TCP-friendly.
- **Scalability:** The performance of the adaptation scheme must not deteriorate with increasing numbers of receivers.

- **Heterogeneity:** The adaptation scheme needs to take into account the heterogeneity of the Internet and must aim at satisfying the requirements of a large part of the receivers if not all possible receivers.

BACKGROUND

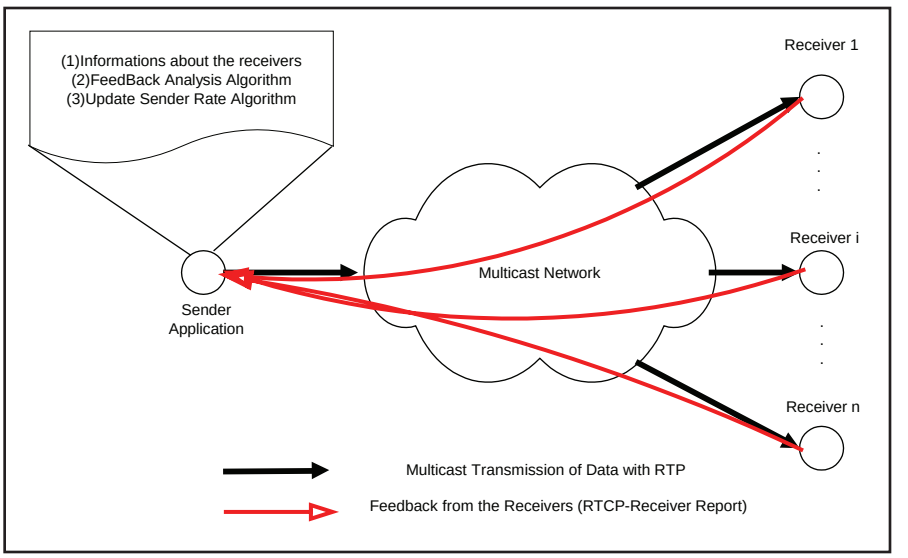
When someone multicasts multimedia data over the Internet, he or she has to accommodate receivers with heterogeneous data reception capabilities. To accommodate heterogeneity, the sender application may transmit one multicast stream and determine the transmission rate that better satisfies most of the receivers, may transmit at multiple multicast streams with different transmission rates and allocate receivers at each stream, or may use layered encoding and transmit each layer to a different multicast stream.

The single multicast stream approach has the disadvantage that clients with a low bandwidth link will always get a high-bandwidth stream if most of the other members are connected via a high bandwidth link and vice versa. The previously described problem can be overcome with the use of a multi-stream multicast approach. Single multicast stream approaches have the advantages of easy encoder and decoder implementation and simple protocol operation, due to the fact that during the single multicast stream approach there is no need for synchronisation of receivers' actions (as is required for the multiple multicast streams and layered encoding approaches).

The methods proposed for the multicast transmission of multimedia data over the Internet can be generally divided in three main categories, depending on the number of multicast streams used:

- The sender uses a single multicast stream for all receivers (Bouras & Gkamas, 2003). This results to the most effective use of the network resources, but on the other hand the fairness problem among the receivers arises, especially when the receivers have very different capabilities. The subject of adaptive multicast of multimedia data over networks with the use of one multicast stream has engaged many researchers. During the adaptive multicast transmission of multimedia data in a single multicast stream, the sender application must select

Figure 1. Architecture of a single stream multicast transmission mechanism



the transmission rate that satisfies most of the receivers with the current network conditions. Three approaches can be found in the literature for the implementation of the adaptation protocol in a single stream multicast mechanism: equation based (Pandhye et al., 1999), network feedback based (Jiang, Ammar, & Zegura, 1998; Sisalem, 1998) or based on a combination of the previous two approaches (Sisalem & Wolisz, 2000a).

- **Simulcast:** The sender transmits versions of the same video, encoded in varying degrees of quality. This results to the creation of a small number of multicast streams with different transmission rates (Bouras, Gkamas, Karaliotas, & Stamos, 2001). The different multicast streams carry the same video information but in each one the video is encoded with different bit rates, and even different video formats. Each receiver joins in the stream that carries the video quality, in terms of transmission rate, that it is capable of receiving. The main disadvantage in this case is that the same multimedia information is replicated over the network but recent research has shown that under some conditions simulcast has better behavior that multicast transmission of layered encoded video (Kim & Ammar, 2001).
- The sender uses *layered encoded* video, which is video that can be reconstructed from a number of discrete data layers, the basic layer, and more additional layers, and transmits each layer into different multicast stream (Legout & Biersack, 2000; Sisalem & Wolisz, 2000b). The basic layer provides the basic quality and the quality improves with each additional layer. The receivers subscribe to one or more multicast streams depending on the available bandwidth into the network path to the source.

SINGLE STREAM MULTICAST TRANSMISSION OF MULTIMEDIA DATA

In such mechanism a sender application transmits multimedia data to a group of n receivers with the use of multicast in one stream. The sender application is using *RTP*/*RTCP* protocols for the transmission of the multimedia data. Receivers receive the multimedia data and inform the sender application for the quality of the transmission with the use of RTCP receiver reports. The sender application collects the RTCP receiver reports, analyses them and determines the transmission rate that satisfy most the group of receivers with the current network conditions.

During the single stream multicast transmission the sender usually runs two algorithms:

- **Feedback analysis algorithm:** Feedback analysis algorithm analyses the feedback information that the receivers sends to the sender application (most mechanisms use RTCP receiver reports for this purpose), concerning the transmission quality of the multimedia data. Every time the sender application receives feedback from a receiver, it runs the feedback analysis algorithm in order to estimate the preferred transmission rate, which will satisfy that receiver. The receiver's preferred transmission rate represents the transmission rate that this receiver will prefer if it was the only one receiver in the multicast transmission of the multimedia data.
- **Update sender rate algorithm:** The sender application in repeated time periods estimates the transmission rate for multicasting the multimedia data with the use of the update sender rate algorithm. The estimation of the sender application transmission rate is aiming to increase

the satisfaction of the group of receivers. When the sender application estimates the new transmission rate, it tries to provide to the group of receivers the best satisfaction that the current network conditions allow.

SIMULCAST

In such a mechanism, the server is unique and responsible of:

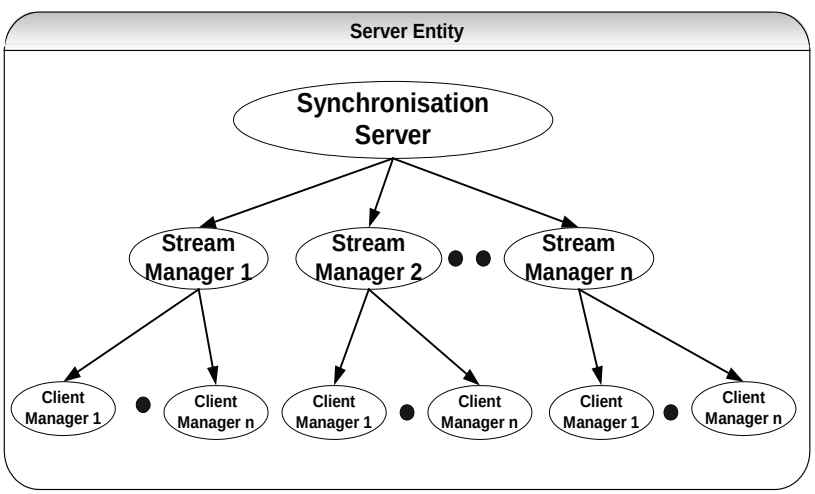
- Creating the n different multicast streams (in most mechanisms a small number of multicast streams, usually 3 or 4 are enough)
- Setting each stream's bandwidth limits
- Tracking if there are any clients that are not handled with fairness
- Providing the mechanisms to the clients to switch streams whenever they consider that they should be in another stream closer to their capabilities

Figure 2 shows the organisation and the architecture of the server entity. The server generates n different stream managers. In each stream manager**,** an arbitrary number of client managers is assigned. Each client manager corresponds to a unique client that has joined the stream controlled by this stream manager. The synchronisation server is responsible for the management, synchronisation, and intercommunication between stream managers.

The stream manager entity is responsible for the maintenance and the monitoring of one of the n different multicast streams. Also the stream manager entity has all the intra-stream adaptation mechanisms for the adjustment of the transmission rate. The stream manager periodically gathers the states reported by all client managers belonging to it at the end of a specific, fixed time period. It then uses an appropriate algorithm that tries to improve fairness between clients by determining whether a lower or a higher bit rate is more appropriate. Whenever a client cannot be satisfied by a stream due to the fact that most of the other clients have much higher or much lower reception capabilities, the stream manager informs it that it has to move to a lower or higher quality stream.

Each client manager corresponds to a unique client (for scalability issues a small representative group of clients may have a corresponding client manager). It processes the RTCP reports generated by the client and can be considered

Figure 2. The architecture and the data flow of the server



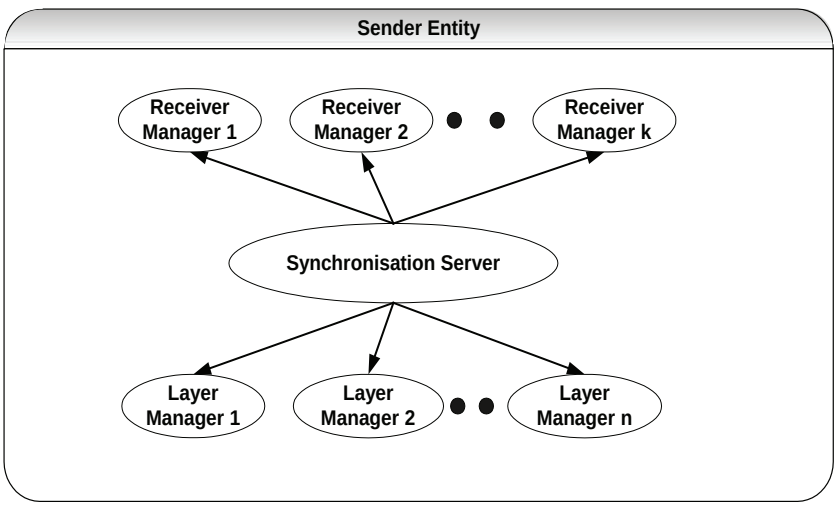
as a representative of the client at the side of the server. It can interact only with one stream manager at a given time, the stream manager controlling the stream from which the client is receiving the video. Client manager receives the RTCP reports from the client and processes them based on packet loss rate and delay jitter information. It then makes an estimation of the state of the client, based on the current and a few previous reports that it stores in a buffer.

LAYERED ENCODING

In such mechanism, the sender transmits multimedia data to a group of m receivers with the use of multicast. The sender is using the layered encoding approach, and transmits the video information in n different layers (the basic layer and n-1 additional layers). The sender transmits each layer into a different RTP/RTCP multicast session. The transmission rate within each layer is adapting within its limits (each layer has an upper and lower limit in its transmission rate) according to the capabilities of the receivers listening up to it. The receivers join the appropriate number of layers which better suit their requirements (available bandwidth between the sender and the receiver, etc) and if during the transmission of multimedia data the network conditions to the path between them and the sender change, the receivers have the capability to receive more or less video layers in order to accomplish better their requirements. The communication between the sender and the receivers is based on RTP/RTCP sessions and the sender is using the RTP protocol to transmit the video layers and the participants (the sender and the receivers) use the RTCP protocol in order to exchange control messages.

Figure 3 shows the organisation and the architecture of the sender entity. The sender generates n different layer managers. Each layer manager is responsible for the transmission of a video layer. The sender creates a new receiver manager every time receives a RTCP report from a new receiver. Each receiver manager corresponds to a unique receiver (for scalability issues a small representative group of receivers may have a corresponding receiver manager). It processes the RTCP reports generated by the receiver and can be considered as a representative of the receiver at the side of the sender. In addition, the synchronisation server is responsible for the management, synchronisation and intercommunication between layer managers

Figure 3. The architecture and the data flow of the sender



and receiver managers. If a receiver manager does not receive RTCP reports from the receiver which it represents for a long time, it stops its operation and releases its resources.

Each receiver measures the characteristics of the path, which connects it with the sender and informs the sender with the use of receiver reports.

EVALUATION PARAMETERS

During the multicast transmission of multimedia data over the Internet the overall target is the optimal usage of the network resources and for this reason an appropriate mechanism is used. In order to evaluate those mechanisms there are the following criteria:

- **Network congestion:** The goal of the multicast transmission mechanisms is to increase the usage of the available bandwidth and decrease the packet losses of all the applications that transmit data in the same network path with the network path of the multicast data.
- **Scalability:** During the multicast transmission of multimedia data, the multimedia data may be received by a large number of receivers. The performance of the selected mechanism must not be downgraded when the number of the receivers of the multicast data is increased. This means that the complexity and the performance of the used mechanism must be acceptable even when a large number of receivers receive the multimedia data through the multicast transmission.
- **Adaptation speed:** With the term adaptation speed we refer to the time needed from the begging of the multicast transmission of the multimedia data until the selected mechanism achieves a stable operation. This time must be relatively small and the performance of the mechanism is better when this time is small.
- **TCP friendliness:** Most of the Internet traffic is TCP traffic. Any application that transmits data over the Internet should have a friendly behaviour toward the other flows that coexist in today's Internet and especially toward the TCP flows that comprise the majority of flows.
- **User satisfaction:** It is difficult to measure the user satisfaction. For example, studies has show that during the transmission of MPEG video, just 3% packet loss can result up to 30% reduction of the presentation quality. As a result the satisfaction of the end user is influenced very much from the packet loss.

TRANSMISSION OF MULTIMEDIA DATA

The transmission of the multimedia data is based on the protocols RTP/RTCP. The protocol RTP is used for the transmission of the multimedia data from the server to the client and the client uses the RTCP protocol, in order to inform the server of the transmission quality.

The RTP/RTCP protocols have been designed for the transmission of real time data like video and audio. Although the RTP/RTCP protocols were initially designed for multicast transmission, they were also used for unicast transmissions. RTP/RTCP can be used for one-way communication like video on demand or for two-way communication like videoconference. RTP/RTCP offers a common platform for the representation of synchronisation information that real time applications need. The RTCP protocol is the control protocol of RTP. The RTP protocol has been designed to operate in cooperation with the RTCP protocol, which provides information about the transmission quality.

RTP is a protocol that offers end to end transport services with real time characteristics over packet switching networks like IP networks. RTP packet headers include information about

the payload type of the data, numbering of the packets and timestamping information.

RTCP offers the following services to applications:

- **QoS monitoring:** This is one of the primary services of RTCP. RTCP provides feedback to applications about the transmission quality. RTCP uses sender reports and receiver reports, which contain useful statistical information like total transmitted packets, packet loss rate and delay jitter during the transmission of the data. This statistical information is very useful, because it can be used for the implementation of congestion control mechanisms.
- **Source identification:** RTCP source description packets can be used for identification of the participants in a RTP session. In addition, source description packets provide general information about the participants in a RTP session. This service of RTCP is useful for multicast conferences with many members.
- **Inter-media synchronisation:** In real time applications it is common to transmit audio and video in different data streams. RTCP provides services like timestamping, which can be used for inter-media synchronisation of different data streams (for example synchronisation of audio and video streams).

More information about RTP/RTCP can be found in RFC 3550 (Schulzrinne, Casner, Frederick, & Jacobson, 2003).

FUTURE TRENDS

The mechanisms described in the previous paragraphs have been proposed for installation and operation over the Internet. One interesting extension of the previous mechanisms is the adaptation of the previous mechanisms to operate over mobile networks. The multicast transmission of multimedia data over mobile networks is a challenge due to the fact the one of the basic characteristics of mobile networks is the continuously changing environment. In order to adapt the previously described mechanisms for usage over mobile networks various issues must be considered such as more efficient encodings.

CONCLUSION

The multicast transmission of real time multimedia data is an important component of many current and future emerging Internet applications such as videoconferencing, distance learning, and video distribution. The heterogeneous nature of the Internet makes the multicast transmission of real time multimedia data a challenge. Different receivers of the same multicast stream may have different processing capabilities, different loss tolerance and different bandwidth available in the paths leading to them.

When multicast multimedia data is transmitted over the Internet, receivers with heterogeneous data reception capabilities have to be accommodated. To accommodate heterogeneity, the sender application may transmit one multicast stream and determine the transmission rate that satisfies most of the receivers, it may transmit at multiple multicast streams with different transmission rates and allocate receivers at each stream or it may use layered encoding and transmit each layer to a different multicast stream.

REFERENCES

Bouras, C., & Gkamas, A. (2003). Multimedia transmission with adaptive QoS based on real time protocols. *International Journal of Communications Systems, 16*(2), 225-248, Wiley InterScience.

Bouras, C., Gkamas, A., Karaliotas, A., & Stamos, K., (2001). Architecture and performance

evaluation for redundant multicast transmission supporting adaptive Qos. *International Conference on Software, Telecommunications, and Computer Networks*, Split, Dubrovnik (Croatia) Ancona, Bari (Italy), October 8-11, 2001, pp. 585-592.

Floyd, S., & Fall, K. (1998). Promoting the use of end-to-end congestion control in the Internet. IEEE/ACM Transactions on Networking, Volume 7, Issue 4 (August 1999), pp. 458-472.

Jiang, T., Ammar, M., & Zegura, E. (1998). Inter-receiver fairness: A novel performance measure for multicast ABR sessions. SIGMETRICS, June 22-26, Madison, Wisconsin, USA, pp. 202-211

Kim, T., & Ammar, M. (2001). A comparison of layering and stream replication video multicast schemes. In *Proceedings of the NOSSDAV'01*, Port Jefferson, NY, June 25-26, 2001, pp. 63-72.

Legout, A., & Biersack, E. (2000), PLM: Fast convergence for cumulative layered multicast transmission schemes. In *Proceedings of ACM SIGMETRICS'2000*, Santa Clara, CA, USA, June 17-21, 2000, pp. 13-22.

Pandhye, J., Kurose, J., Towsley, D., & Koodli, R, (1999). A model based TCP-friendly rate control protocol. In *Proceedings of the International Workshop on Network and Operating System Support for Digital Audio and Video*, Basking Ridge, NJ

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). RTP: A transport protocol for real-time applications, RFC 3550, IETF.

Sisalem, D. (1998). Fairness of adaptive multimedia applications. IEEE International Conference on Communications. Conference Record. Affiliated with SUPERCOMM'98 IEEE,. p.891-5 vol.2. 3 vol. xxxvii+1838 pp

Sisalem, D., & Wolisz, A. (2000a). LDA+ TCP-friendly adaptation: A measurement and comparison study. The *10th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Chapel Hill, NC, USA, June 26-28 2000, pp. 1619-1622.

Sisalem, D., & Wolisz, A. (2000b). MLDA: A TCP-friendly congestion control framework for heterogeneous multicast environments. The *8th International Workshop on Quality of Service*, Pittsburgh, PA, June 5-7, 2000, pp. 65-74.

KEY TERMS

Layered Encoding: Transmission of the multimedia data in n different layers the basic layer and n-1 additional layers.

Multicast: Transmitting data simultaneously to many receivers without the need to replicate the data.

Multimedia Data: Multimedia data refers to data that consist of various media types like text, audio, video, and animation.

Quality of Service (QoS): Quality of service refers to the capability of a network to provide better service to selected network traffic.

RTP/RTCP: Protocol that is used for the transmission of multimedia data. The RTP performs the actual transmission and the RTCP is the control and monitoring transmission.

Simulcast: Transmission of the same multimedia data in multiple multicast streams with different transmission rates.

This work was previously published in Encyclopedia of Internet Technologies and Applications, edited by M. Freire and M. Pereira, pp. 316-322, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.13
IP Multimedia Subsystem (IMS) for Emerging All-IP Networks

Muhammad Sher
Technical University of Berlin, Germany

Fabricio Carvalho de Gouveia
Technical University of Berlin, Germany

Thomas Magedanz
Technical University of Berlin, Germany

INTRODUCTION AND BACKGROUND

Today the traditional telecommunication technology is declining because of popularity and increasing demand of voice over IP (VoIP) due to the reason that deployment, maintenance, and operation of data networks based on IP infrastructure are less costly than the voice networks. Consequently, it is straightforward to think relaying all types of communications on data networks rather than maintaining in parallel two network technologies. On the other hand, today we see increasing demand of *integrated multimedia services*, bringing together Internet applications with telecommunications. In prospect of these global trends, the mobile communications world has defined an all-IP network vision within the evolution of cellular systems, which integrates cellular networks and Internet. This is the IP multimedia system (IMS) (3GPP, TS 23.228, 2005), namely overlay architecture for provisioning of multimedia services such as VoIP and videoconferencing on top of globally emerging 3G broadband packet networks.

In the face of the IP network vision, namely the use of fixed and mobile IP networks for both data and voice/multimedia information looking for the target service control architecture. The IMS is an approach to provide overlay service delivery platform (SDP) (Magedanz & Sher, 2006) architecture for IP networks, entirely built on Internet protocols defined by the Internet engineering task force (IETF), which have been extended on request of 3GPP (Third Generation Partnership Project) to support telecommunications require-

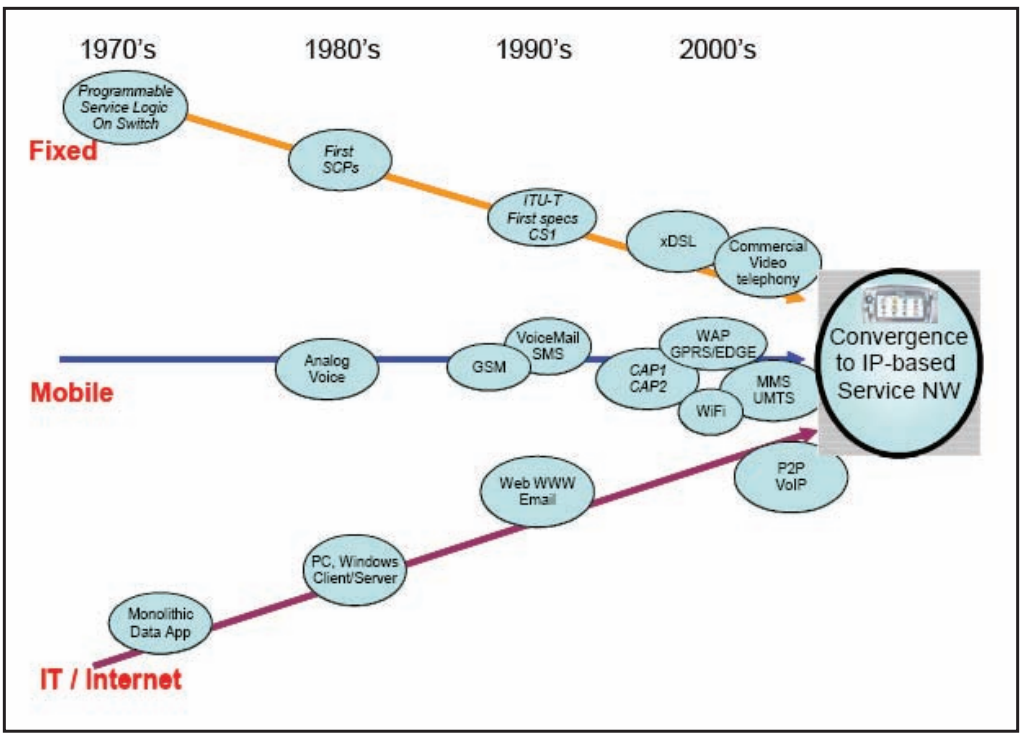
ments such as security, accountability, quality of service, etc. Mobile operators today face the problem that mobile users can gain access to the Internet and make use of Internet services such as *instant messaging*, *presence*, chat, content download, etc. On the other hand, the operators define a minimum SDP architecture for providing QoS, security, and charging for IP-based services, while providing maximum flexibility for the realization of value added and content services.

EVOLUTION TOWARD CONVERGED NETWORKS

The communication era is passing through the evolutionary phase of fixed-mobile convergence (FMC) and *voice-data integration* as shown in Figure 1, and IMS is considered a common platform for FMC granting convergence and compatibility between fixed/mobile networks, multimedia, and converged services support, and providing key community service enablers such as group management, *presence*, IM/PoC, and generic VoIP/MMoIP support (Magedanz, 2005).

In face of such convergence, the need for universal service delivery platforms (SDPs) supporting integrated services emerged. This means that SDP should, in principle, enable the rapid and uniform programming and provision of seamless multimedia services on top of any network environment. There is no doubt, however, that today two main trends are of pivotal importance for SDPs design, namely the support of mobile users and the support of (mobile) multi media data services.

Figure 1. Toward fixed-mobile-internet convergence



IMS MOTIVATION AND STANDARDIZATION

IMS is designed to provide fancy and attractive Internet services everywhere using mobile networks based on IP protocols and standards with emphasis on QoS, dynamic charging, and integration of different services and roaming facility on reasonable service charges. The IMS provides easy and efficient ways to integrate different services, even from third parties, and enables the *seamless integration* of legacy services and is designed for consistent interactions with circuit switched domains. The IMS manages event-oriented quality of service policies (e.g., use of VoIP and HTTP in a single session—VoIP has QoS, HTTP is best effort). These systems (IMSs) also have event-oriented charging mechanism policies--means change specific events on the appropriate level. If two events need the same IP resources we may charge them differentially for the same user in the same session (Poikselkae, Mayer, Khartabil, & Niemi, 2004). These characteristics make IMS the future technology in a comprehensive service and application-oriented network.

The IMS has been standardized since the beginning of this century within Release 5 and extended in Release 6 within 3GPP and 3GPP2 (Third Generation Partnership Project 2) focus to UMTS/CDMA2000 data packet networks. The Release 5 standards have been driven by the vision to define the IMS for providing multimedia services including VoIP. The IMS is supposed to be standardized access-independent IP-based architecture that interworks with existing voice and data networks for fixed (e.g., PSTN, ISDN) Internet and mobile users (e.g., GSM, CDMA). The IMS architecture makes it possible to establish peer-to-peer IP communications with all types of clients with quality of services and complete service delivery functionalities.

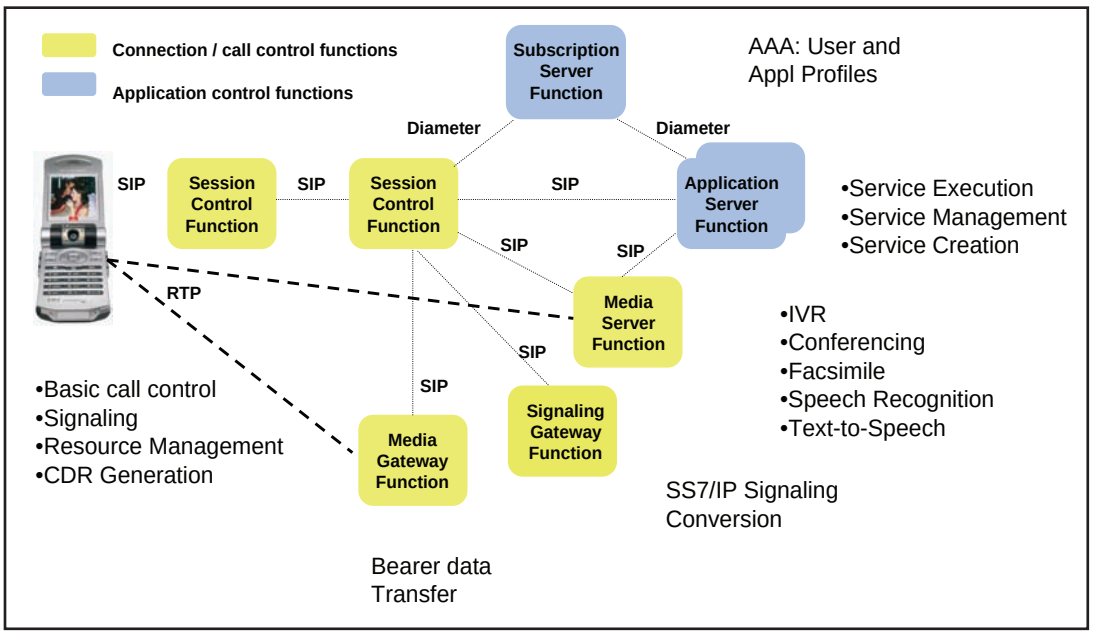
IMS Release 6 is fixing the shortcomings in Release 5 and also contains novel features like *presence*, messaging, conferencing, group managements and local services. Release 6 has optimized the IMS to provide the envisaged IMS killer application push-to-talk (over cellular). The IMS has been planned for deployment in 3G wireless networks around 2006. It also provides additional security features like confidentiality protection of *SIP* messages, usage of public key infrastructure, and subscriber certificates.

In addition, since 2004 ETSI *TISPAN* (Telecommunication and Internet converged Services and Protocols for Advanced Networking) is looking at service infrastructures for fixed-mobile convergence and next generation networks, which extends the IMS to make it applicable on top of various access networks (i.e., WLANs and particular for fixed Internet (DSL)). The recent IMS Release 7 is a joint cooperation work of 3GPP and TISPAN addressing all IP networks infrastructure.

IMS ARCHITECTURE AND KEY COMPONENTS

The IMS defines service provision architecture and it can be considered as the next generation service delivery platform framework. It consists of modular design with open interfaces and enables the flexibility for providing multimedia services over IP technology. IMS does not standardize specific services but uses standard service enablers (e.g., *presence* inherently supports multimedia over IP and VoIP) (Magedanz et al., 2006). In IMS architecture, SIP protocol (IETF, RFC 3261, 2002) is used as the standard signaling protocol that establishes controls, modifies, and terminates voice, video, and messaging sessions between two or more participants. The related signaling servers in the architecture are referred to as call state control functions (CSCFs) and distinguished by their specific functionalities. It is important to note that an IMS compliant end user system has to provide the necessary IMS protocol support, namely SIP, and the service related media codecs

Figure 2. IMS high level functional diagram



for the multimedia applications in addition to the basic connectivity support (e.g., GPRS, WLAN, etc.). Figure 2 displays a generic IP-based IMS functional diagram.

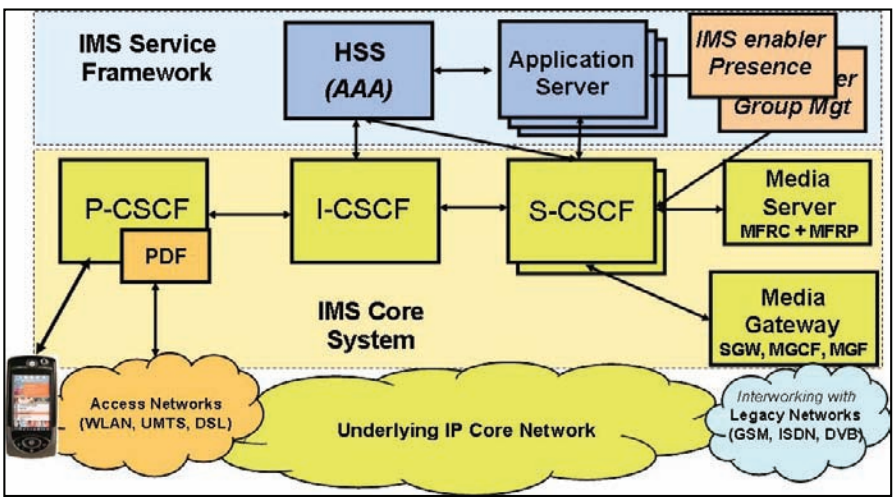
The functionality related to authentication, authorization, and accounting (AAA) within the IMS is based on the IETF *diameter* protocol (IETF, RFC 3588, 2003) and is implemented in the home subscriber system (HSS), CSCFs, and various other IMS components in order to allow charging functionality within the IMS. Instead of developing the protocol from scratch, diameter is based on the *remote authentication dial in user service* (RADIUS), which has previously been used to provide AAA services for dial-up and terminal server across environments. The other protocol, which is important for multimedia contents, is *real-time transport protocol* (RTP), which provides end-to-end delivery for real-time data. It also contains end-to-end delivery services like payload-type (codec) identification, sequence numbering, time stamping, and delivering monitoring for real-time data. RTP provides QoS monitoring using the RTP control protocol (RTCP), which conveys information about media session participants.

The IMS entities and key functionalities can be classified in six categories--session management and routing family (CSCFs), databases (HSS, SLF), interworking elements (BGCF, MGCF etc.), services (application server, MRCF, MRFP), support entities (THIG, SEG, PDF) and charging (Poikselkae et al., 2004). Now we describe the most important components and parts of IMS architecture:

- **Proxy Call State Control Function (P-CSCF):** It is the first contact point within IP multimedia core network subsystem. Its address was discovered by UEs following packet data protocol (PDP) context activation. The P-CSCF behaves like a proxy accepting requests and services them internally or forwards them. It performs functions like authorizing the bearer resources for the appropriate QoS level, emergency calls, monitoring, header (de)compression, and identification of I-CSCF.

- **Interrogating Call State Control Function (I-CSCF):** It is the contact point within an operator's network for all connections destined to a subscriber of that network operator, or a roaming subscriber currently located within that network operator's service area. There may be multiple I-CSCFs within an operator's network. I-CSCF performs functions like assigning an S-CSCF to a user performing SIP registration/charging and resource utilization (i.e., generation of charging data records (CDRs)/acting as a topology hiding inter-working gateway (THIG)).
- **Serving Call State Control Function (S-CSCF):** It performs session control services for the endpoint and maintains session state as needed by network operator for support of services. Within an operator's network, different S-CSCFs may have different functionality. The important functions performed by S-CSCF includes user registration/interaction with services platforms for the support of services. The S-CSCF decides whether an application server is required to receive information related to incoming SIP session request to ensure appropriate service handling. The decision at the S-CSCF is based on filter information received from the HSS. This filter information is stored and conveyed on a per application server basis for each user.
- **Home Subscriber Server:** HSS is equivalent to HLR (home location register) in 2G systems but is extended with two diameter-based reference points. It is the master database of an IMS that stores IMS user profiles including individual filtering information, user status information, and application server profiles.
- **Application Servers:** It provides service platform in IMS environment. It does not address how multimedia/value added applications are programmed but only well defined signalling and administration interfaces (ISC and Sh), and SIP and Diameter protocols are supported. This enable developers to use almost any programming paradigm within a SIP AS such as legacy intelligent network servers (i.e., CAMEL support environments), OSA/Parlay servers/gateways, or any proven VoIP SIP programming paradigm like SIP Servlets, call programming language (CPL) and common

Figure 3. IMS components and architecture



gateway interface (CGI) scripts, etc. The SIP AS is triggered by the S-CSCF, which redirects certain sessions to the SIP AS based on the downloaded filter criteria or by requesting filter information from HSS in a user-based paradigm. The SIP AS itself comprises filter rules to decide which of the applications deployed on the server should be selected for handling the session. During execution of service logic it is also possible for SIP AS to communicate with HSS to get additional information about a subscriber or to be notified about changes in subscriber profile.

- **Media Processing:** The media resource function (MRF) can be split up into media resource function controller (MRFC) and media resource function processor (MRFP). It provides media stream processing resources like media mixing, media announcements, media analysis, and media transcoding as well speech (Poikselkae et al., 2004). The other three components are border gateway control function (BGCF), media gate control function (MGCF) and Media gate (MG), which perform bearer interworking between RTP/IP and bearers used in the legacy networks.
- **IMS End User System:** It is important to note that an IMS compliant end user system has to provide the necessary IMS protocol support, namely SIP, and service related media codecs for multimedia applications in addition to the basic connectivity support (e.g., GPRS, WLAN, etc.).

IMS FEATURES AND SERVICES

The IMS is designed to provide number of key capabilities required to enable new IP services via mobile and fixed networks. The important key functionalities, which enable new mobile IP services, are:

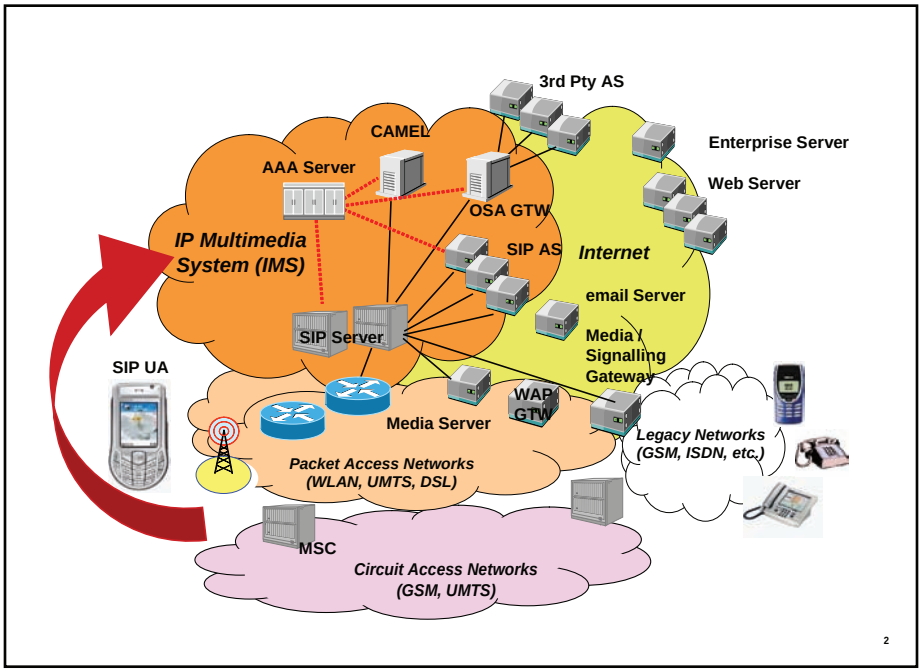
- Multimedia session negotiation and management
- Quality of service management
- Mobility management
- Service execution, control, and interaction

The IMS services are assumed to be addressed by open mobile alliance (OMA), which was created by WAP Forum in June 2002 with open mobile architecture objective and consists of more than ten mobile industries. The OMA SIP-based service enablers are specified on top of IMS as common platform (e.g., presence and group management etc.) as shown in Figure 4.

The important services provided by IMS are:

- **Push-to-talk (PTT) over cellular (PoC):** The PoC is *1-to-n* half-duplex communication and the key PTT functions include presence, group list management, PTT media processing, and PTT application logic (including floor control handling). These are bundled tightly together in the vendor-specific PoC deployments, but from 2006 onwards, IMS-based PTT implementation will be deployed. The idea is to enable the reuse of PTT core ingredients for other service offers, such as presence-based services. PoC content are short, instructional, and immediate.
- **Multimedia conferencing and group chat:** It is real time communication providing functionality about reservation and scheduling, document sharing and whiteboard facility, additional group or 1-to-1 chat, and instant messaging availability.
- **Click to dial:** When a user clicks on the screen button, IMS will negotiate and eventually set up automatically a voice session with one or more other users.
- **Dynamic push service:** Allow users to receive disable information based on many factors (e.g., preference, presence, geographic location, device type, and capabilities).

Figure 4. IMS as multimedia service enabler



OPEN ISSUES AND FUTURE TRENDS

The new multimedia services are a demanding combination of service capability features. Most likely upcoming services will also rely on features like presence, group-list management, additional logic, and other features on operator network (e.g., location, SMS, MMS). It is obvious that service capability features must be reused for scalability and capital expenses reasons. The future trends are:

- How to manage and orchestrate services
- How to create stringent services that bundle service capability features
- How to operate the network for services in a secure way

As mentioned in 3GPP specifications, the adoption of OSA/Parlay (parlay open service architecture) concepts and technologies can contribute a lot. OSA/Parlay already provides an industry standard that enables unified access with gateway character to service capability features of operators' network. Even secure access by third parties can be handled by the OSA/Parlay framework. This framework may control resources by assuming there is secure access for third parties network.

Recently, there are many IMS pre-products originating from VoIP and wireless telecommunications market. But, there is not yet any commercial IMS deployment within operator networks. However, first push to talk (PTT) service implementations mushrooming around the globe can be regarded as the first big trials for IMS technologies. However, there are still many open issues within IMS architecture and the 3GPP IMS standardization is ongoing, particular in the field of applying the IMS on top of different access networks (i.e., WLAN, WIMAX, and DSL) and IMS evolution toward an all-IP network.

CONCLUSION

The IP multimedia subsystem (IMS) defined by the 3rd Generation Partnership Projects (3GPP) and multimedia domain (MMD) by 3GPP2 is today considered the global service delivery platform (SDP) standard for providing multimedia applications in next generation networks (NGN). IMS defines an overlay service architecture that merges the paradigms and technologies of the Internet with the cellular and fixed telecommunication worlds. Its architecture enables the efficient provision of an open set of potentially highly integrated multimedia services, combining Web browsing, e-mail, instant messaging, presence, VoIP, video conferencing, application sharing, telephony, unified messaging, multimedia content delivery, etc. on top of possibly different network technologies. As such IMS enables various business models for providing seamless business and consumer multimedia applications. IMS platform also supports interworking with Internet and legacy networks.

In this article, we have provided background, motivation, standards, services, and architecture of IP multimedia subsystem (IMS), which is considered the next generation service platform toward all-IP networks. We have also tried to highlight the future trends and open issues in IMS.

REFERENCES

3GPP, Third Generation Partnership Project (3GPP). www.3gpp.org.

3GPP (2005), IP Multimedia Subsystems (IMS); Stage 2, TS 23.228, Version 7.2.0, *Third Generation Partnership Project.* Retrieved December 15, 2005, from www.3gpp.org

3GPP2, Third Generation Partnership Project 2 (3GPP2). www.3gpp2.org.

Calhoun, P., Loughney, J., Guttman, E., Zorn, G., & Arkko, J. (2003). RFC 3588 Diameter Base Protocol. *Internet Engineering Task Force* (*IETF*). Retrieved November 28, 2005, from http://www.ietf.org/rfc3588.txt?number=3588

ETSI TISPAN, Telecommunications and Internet converged Services and Protocols for Advanced Networking. http://portal.etsi.org/tispan/TISPAN_ToR.asp.

IETF, RFC 3261 (2002), Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., & Schooler, E., SIP: Session Initiation Protocol.

IETF, RFC 3588 (2003), Calhoun, P., Loughney, J., Guttman, E., Zorn, G., & Arkko, J., Diameter Base Protocol.

Magedanz, T. (2005). Tutorial IEEE ISCC. *IEEE Symposium on Computer and Communications.*

Magedanz, T., & Sher, M. (2006). IT-based open service delivery platforms for mobile networks--From CAMEL to the IP multimedia system. In P. Bellavista & A. Corradi (Eds.), *Mobile middleware* (pp. 1001-1037), Taylor & Francis CRC Press, Florida, USA.

OMA, Open Mobile Alliance, http://www.openmobilealliance.org

OSA/Parlay "Parlay Open Service Architecture", http://www.parlay.org/en/index.asp

Poikselkae, M., Mayer, G., Khartabil, H., & Niemi, A. (2004). *The IMS: IP multimedia concepts and services in the mobile domain*. West Sussex: John Willey & Sons Ltd.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). RFC 3261 SIP: Session Initiation Protocol. *Internet Engineering Task Force* (*IETF*). Retrieved November 26, 2005, from http://www.ietf.org/rfc/rfc3261.txt?number=3261

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). RFC 3550 A Transport Protocol for Real-Time Applications. *Internet Engineering Task Force* (*IETF*). Retrieved November 28, 2005, from http://www.ietf.org/rfc/rfc3588.txt?number=3588

KEY TERMS

Fixed-Mobile Convergence (FMC): FMC is the merger of fixed and mobile communication paradigms and IMS is considered as common platform for FMC granting convergence and compatibility between fixed/mobile networks, multimedia and converged services support, and providing key community service enablers such as group management, presence, IM/PoC and generic VoIP /MMoIP support.

IP Multimedia Subsystem (IMS): The IP multimedia subsystem (IMS) is standardised by 3GPP & 3GPP2 as next generation network (NGN) architecture to provide mobile and fixed multimedia services. It is based on VoIP, SIP, and IP protocols on the recommendation on 3rd Generation Partnership Project on top of GPRS/UMTS.

Next Generation Network (NGN): A next generation network (NGN) as defined by ITU-T is a packet-based network able to provide services including telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It offers unrestricted access by users to different service providers. It supports generalized mobility, which will allow consistent and ubiquitous provision of services to users.

Open System Architecture (OSA): The open system architecture (OSA) describes the service architecture for 3rd generation mobile telecommunications network or universal mobile telecommunication networks (UMTS). OSA Standards are being developed and published by as part of 3rd Generation Partnership Project (3GPP) and ETSI. The OSA APIs are developed by parlay group and is called OSA/Parlay service framework.

Service Delivery Platform (SDP): SDP is a foundation for the creation, deployment, provision, control, charging, and management of telecommunication services provided to the end users. The SDPs represent the programming interface enabling programming of the underlying network capabilities and therefore are primarily based in the usage of information technologies (ITs). SDP may apply to many technologies including VoIP, IPTV, HSD (high speed data), Mobile Telephony, Online Gaming, etc.

Third Generation Partnership (3GPP): The 3rd Generation Partnership Project (3GPP) is a collaboration agreement between mobile industries that was established in December 1998 to extend the GSM specifications toward Third Generation (3G) mobile system specifications within the scope of the ITU's IMT-2000 project. 3GPP specifications are generally known as the UMTS system.

Voice Over IP (VoIP): Voice over Internet protocol (VoIP) is the technology for sending voice calls on IP-based network instead of using conventional voice telecommunication networks. VoIP protocol is used to carry voice signals over IP networks.

This work was previously published in Encyclopedia of Internet Technologies and Applications, edited by M. Freire and M. Pereira, pp. 249-256, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Index

Symbols

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z